Knowledge-Based Systems 54 (2013) 199–206

Contents lists available at ScienceDirect

Knowledge-Based Systems

journal homepage: www.elsevier.com/locate/knosys

Discriminative sparse coding on multi-manifolds

Jim Jing-Yan Wang [a], Halima Bensmail [b], Nan Yao [c,d], Xin Gao [a,e,*]

[a] Computer, Electrical and Mathematical Sciences and Engineering Division, King Abdullah University of Science and Technology (KAUST), Thuwal 23955-6900, Saudi Arabia
[b] Qatar Computing Research Institute, Doha 5825, Qatar
[c] Department of Instrumentation Engineering, Shanghai Jiao Tong University, Shanghai 200240, China
[d] National Key Laboratory for Novel Software Technology, Nanjing University, Nanjing 210023, China
[e] Computational Bioscience Research Center, King Abdullah University of Science and Technology (KAUST), Thuwal 23955-6900, Saudi Arabia

ARTICLE INFO

Article history:
Received 21 July 2012
Received in revised form 4 September 2013
Accepted 5 September 2013
Available online 26 September 2013

Keywords:
Data representation
Sparse coding
Multi-manifolds
Large margins

ABSTRACT

Sparse coding has been popularly used as an effective data representation method in various applications, such as computer vision, medical imaging and bioinformatics. However, the conventional sparse coding algorithms and their manifold-regularized variants (graph sparse coding and Laplacian sparse coding), learn codebooks and codes in an unsupervised manner and neglect class information that is available in the training set. To address this problem, we propose a novel discriminative sparse coding method based on multi-manifolds, that learns discriminative class-conditioned codebooks and sparse codes from both data feature spaces 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-conditioned codebooks and codes to maximize the manifold margins of different classes. Lastly, we present a data sample-manifold matching-based strategy to classify the unlabeled data samples. Experimental results on somatic mutations identification and breast tumor classification based on ultrasonic images demonstrate the efficacy of the proposed data representation and classification approach.

© 2013 The Authors. Published by Elsevier B.V. Open access under CC BY-NC-SA license.

1. Introduction

Sparse coding (Sc) [1–4] has been successfully applied in many pattern recognition applications as a part-based data representation method, including face recognition [5], speech recognition [6], handwritten digit recognition [7], image clustering [7], etc. Given a set of data feature vectors organized as an input data matrix, Sc aims to find a basis vector pool (also known as the codebook), 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 samples to be encoded might be ignored [8,9]. As a result, similar data vectors may be represented as totally different sparse codes based on such codebooks, bringing instability to the Sc and harming the robustness of sparse coding-based pattern recognition applications [10,11]. To overcome this disadvantage, Graph-regularized Sparse Coding (GraphSc) and Laplacian Sparse coding (LSc) were proposed by Zheng et al. [7] and Gao et al. [10,11], respectively. In both 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 codes change smoothly along the geodesics of the data manifold [7,10,11]. Moreover, Sc using Manifold Projections (ScMP) was proposed by Ramamurthy et al. [12] to ensure that the data recovered using sparse representation is close to its manifold, by performing regularization using examples from the data manifold.

For most pattern recognition tasks, such as somatic mutations identification [13] and breast tumor classification [14], 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 LapSc and GraphSc are both unsupervised algorithms, meaning that they do not utilize class labels and that they ignore discriminative information that is contained in the labels. Moreover, both GraphSc and LapSc assume that the data samples from different classes define a single general manifold in the feature space and seek common codebooks and coding strategies for all data samples such that the nearby samples are likely to have similar codes. However, as argued by Lu et al. [15,16], "it is still unknown that whether a single manifold could

* Corresponding author at: Computer, Electrical and Mathematical Sciences and Engineering Division, King Abdullah University of Science and Technology (KAUST), Thuwal 23955-6900, Saudi Arabia. Tel.: +966 2 808 0323.
E-mail address: xin.gao@kaust.edu.sa (X. Gao).

0950-7051 © 2013 The Authors. Published by Elsevier B.V. Open access under CC BY-NC-SA license.
http://dx.doi.org/10.1016/j.knosys.2013.09.004

well model the data and guarantee the best recognition accuracy". This assumption is usually not the most suitable.

To solve the problems mentioned above, we assume that the optimal codebooks and coding strategies for different classes should be different due to the intrinsic differences between classes. We propose a novel supervised sparse coding method that learns discriminative codes from both the data features and the class labels. We model the data samples from each class as a manifold such that we can learn optimal codebooks and codes for each class. First, we partition the entire data set into several class-conditioned subsets according to the labels, and then we assume that each subset lies on a class-conditioned manifold, which should be spanned by an independent class-conditioned codebook. Instead of regularizing 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 sparse coding as a class-conditioned data feature reconstruction and a manifold–manifold matching problem and learn multiple codebooks and codes to maximize the manifold margins between different classes. Lastly, we present a data sample-manifold matching-based strategy to classify the test data samples. The proposed algorithm is the first algorithm that divides training sets by the class distributions and then learns the sparse codes respectively. Experimental results on breast tumor classification from ultrasonic images [14] and somatic mutation identification [13] tasks demonstrate the efficacy of the proposed approach.

2. Discriminative sparse coding on multi-manifold (DisScMM)

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

2.1. Object function

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

2.1.1. The sparse reconstruction loss term

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

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

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

2.1.2. The large margin term

Given a sample $x_i \in \mathcal{X}_l$ belonging to the lth 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 samples from the same class as x_i, while inter-class neighbors are the p nearest data samples from different classes from x_i. Using a Gaussian kernel, we first define the class-conditioned 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 it neighbors in $\mathcal{N}_i^{intra}$ as well as that between $x_i \in \mathcal{X}_l$ and $\mathcal{N}_i^{inter}$, respectively,

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

From a classification sample of view, the intra-class variance should be minimized while the inter-class separability should be maximized in the spares coding spaces, such that the class margins can be maximized for sparse coding. To this end, the large margin term of sparse coding is formulated as the following optimization problem for the lth class:

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

On the one hand, the first term of the objective function in (3) is to ensure that if x_i and x_j are close to each other and from the same class, their class-conditioned sparse code representations, v_{li} and v_{lj}, are close as well. On the other hand, the second term of the objective function in (3) ensures that if x_i and x_j are close to each other and from different classes, their class-conditioned sparse code representations, v_{li} and v_{lj}, are separated as far apart as possible. We should notice that, although U_l and V_l are learned separately for each class, they are indirectly connected to maximize the margins between different classes by introducing this margin term.

2.1.3. Objective function of DisScMM

To construct the object function, we first construct the class-conditioned manifold by including the intra- and inter-class neighbors of data samples, $x_i \in \mathcal{X}_l$, 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 samples in this manifold of the lth class are organized as a data matrix, $X_l = [x_n] \in \mathbb{R}^{D\times N_l},\ n = 1, \ldots, N_l,\ x_n \in \mathcal{M}_l$, where $N_l = |\mathcal{M}_l|$ is the number of data samples 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 sample x_n. Then, with the two objective function terms defined above, we have the objective function of DisScMM by combining them as

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

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 matrices whose entries

are $D_{lnn}^{intra}=\sum_{m=1}^{N_l}W_{lnm}^{intra}$ and $D_{lnn}^{inter}=\sum_{m=1}^{N_l}W_{lnm}^{inter}$ separately, $L_l=L_l^{intra}-L_l^{inter}$, and β is the trade-off parameter.

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

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

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

2.2. Optimization

The optimal U_l and V_l of (5) can be solved by the following iterative optimization method introduced in GraphSc [7] and LapSc [10,11]. We adopt the alternate optimization strategy to optimize U_l and V_l in an iterative algorithm. 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 codebook U_l

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

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

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

$$U_l^*=X_lV_l^\top\left(V_lV_l^\top+diag(\lambda^*)\right)^{-1},\tag{7}$$

where $\lambda=[\lambda_1,\ldots,\lambda_K]^\top$,$\lambda_k$ is the Lagrange multiplier [17–20] associated with the kth inequality constraint, $\|u_{lk}\|^2\leqslant c$, and λ^* is the optimal solution of λ. For more details, we refer readers to [1,7].

2.2.2. On optimizing sparse codes V_l

By fixing U_l, the optimization problem (5) becomes

$$\min_{V_l}\|X_l-U_lV_l\|^2+\alpha\sum_{n=1}^{N_l}\|v_{ln}\|_1+\beta Tr\left(V_lL_lV_l^\top\right).\tag{8}$$

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 [21–24] of the multi-manifolds can be rewritten as $Tr(V_lL_lV_l^\top)=\sum_{n,m=1}^{N_l}L_{nm}v_{ln}^\top v_{lm}$. Then, (8) is further reduced to

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

This problem can be optimized by the graph-regularized sparse codes learning algorithm introduced in [7], or by the feature-sign search algorithm introduced in [11]. Here, we adopt the one introduced in [7]. In fact, these two algorithms are basically the same except for the initialization step. Moreover, graph-regularized sparse codes learning introduced in [7] requires the graph weight matrix to be symmetric, while the other one does not.

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

Algorithm 1. The learning procedure of the DisScMM Algorithm.

INPUT: Training sets, $\mathcal{M}_1,\ldots,\mathcal{M}_L$, of L classes of multi-manifolds;
for $l=1,\ldots,L$ **do**
  Construct discriminate graph weight matrices as in (2) and the corresponding Laplacian matrices, L_l, for the lth manifold.
  Initialize the class-conditioned codebook U_l^0 and sparse codes, V_l^0, for lth manifold, by performing Sc on M_l.
  for $t=1,\ldots,t$ **do**
    for $n=1,\ldots,n_l$ **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 the lth manifold.
    end for
    Update the codebook U_l^t while fixing V_l^t by (7) for the lth manifold.
  end for
end for
OUTPUT: The final class-conditioned codebooks U_l^T and sparse codes V_l^T, $l=1,\ldots,L$.

2.3. Classifier of DisScMM

In contrast to traditional Sc methods, which can only be used to represent the date samples, DisScMM can also make use of the discriminative nature of sparse coding on multi-manifolds to perform classification. When a new data sample, x_t, comes in, we match it to all the manifolds and then assign it to the class with the minimum matching error. Assuming that x_t belongs to the lth class, we first calculate its intra-class nearest neighbors, $\mathcal{N}_{lt}^{intra}$, and its inter-class nearest neighbors, $\mathcal{N}_{lt}^{inter}$, from $\mathcal{M}_l$. We also assume that the input of this new data sample has no effect on the discriminate graphs in the sparse codes of $\mathcal{M}_l$, such 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,

$$\mathcal{E}_l(x_t)=\min_{v_{lt}}\left\{\|x_t-U_lv_{lt}\|^2+\alpha\|v_{lt}\|_1+\frac{\beta}{2}\sum_{n:x_n\in\mathcal{N}_{lt}^{inter}}\|v_{lt}-v_{ln}\|^2W_{ltn}^{intra}-\frac{\beta}{2}\sum_{n:x_n\in\mathcal{N}_{lt}^{intra}}\|v_{lt}-v_{ln}\|^2W_{ltn}^{inter},\right\}\tag{10}$$

where W_{ltn}^{intra} and W_{ltn}^{inter} are the intra- and inter-similarities of x_t to the nth data sample of $\mathcal{M}_l$, which is calculated by (2). This optimization problem can also be solved by the algorithm proposed in [7]. We finally assign a label, y_t, to x_t as follows:

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

The classification procedure is summarized in Algorithm 2.

Algorithm 2. The classification procedure of the DisScMM Algorithm.

INPUT: Training sets, $\mathcal{M}_1,\ldots,\mathcal{M}_L$, of L classes of multi-manifolds;
INPUT: The class-conditioned codebooks, U_l, and sparse codes, V_l, for L manifolds, $l=1,\ldots,L$.
INPUT: The input unlabeled data sample, x_t.

(continued on next page)

for $l = 1, \ldots, L$ **do**
Extend the discriminate graph weight matrices by adding x_t as in (2) and compute the corresponding Laplacian matrices, L_l, for the lth 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 the minimum matching error as in (11).
OUTPUT: The class label l^* of x_t.

2.4. Computational complexity

In this section, we discuss the computational complexity of different sparse coding methods. We assume that for all the sparse coding methods, the iterations are repeated T times. In the unsupervised sparse coding methods, including Sc, GraphSc and LapSc, in each iteration, N sparse codes and a dictionary will be learned. The computational complexity is thus $O(TN)$. In our proposed DisScMM algorithm, we learn a dictionary and N_l sparse codes for each class. The computational complexity is thus $O\left(LT\sum_{l=1}^{L}N_l\right)$. Since the equation $\sum_{l=1}^{L}N_l \approx N$ usually holds, the computational complexity can further be reduced to $O(LTN)$. Compared with the unsupervised sparse coding methods, the computational complexity of our supervised DisScMM is almost L times higher. However, we should also note that the learning procedures of DisScMM for each class can be parallelized since they are independent from each other. Assuming that they are fully parallelized, the computational complexity for the lth class will be $O(TN_l)$, which is much lower than that for the unsupervised sparse coding methods.

3. Experiments

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

3.1. Experiment I: Identifying somatic mutations

Profiling tumors for single nucleotide variant (SNV) somatic mutations using next-generation sequencing technology (NGS) plays an important role in the study of cancer genomes [13,25–28]. In this experiment, we evaluate DisScMM on the task of inferring somatic mutations from paired tumor/normal NGS data.

3.1.1. Dataset and setup

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

Training Set The training dataset is selected from exome capture data containing 3369 variants which were predicted using only allelic counts and liberal thresholds [13]. Further resequencing experiments validated 1015 somatic mutations, 471 germ line and 1883 wild-type positions in the dataset. Our selected training dataset contains 800 somatic mutations, and 1800 non-somatic mutations (germ line and wild type).

Test Set The test dataset is selected from a whole genome shotgun dataset containing 113 somatic mutations, 57 germ line mutations and 337 wild types positions [13]. 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 selected 90 somatic mutations and 300 non-somatic mutations to construct the test set.

The ith candidate mutation site of the genome in the dataset is represented by a feature vector, x_i, with 106 feature components constructed from both the tumor and normal data as in [13]. The somatic mutation identifying problem is to predict the label, y_i, of the feature represented site, where y_i is defined as

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

To predict the class labels in the test set, we first learned the codebooks for the somatic mutation manifold and non-somatic mutation manifold using the training set for DisScMM. For this learning procedure, we applied a *10-fold cross-validation* analysis to find the optimal hyper-parameters. Then, the learned DisScMM model was 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 [7] and LapSc [11] and ScMP [12], were also tested as mutation representation methods.

To evaluate the performances of the classification results, we employed recall, precision [29], accuracy, F-score, and Matthews correlation coefficient (MCC) as metrics. Recall, precision and accuracy are defined as

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

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

$$F\text{-}score = 2 \times \frac{Recall \times Precision}{Recall + Precision}. \quad (14)$$

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 a perfect inverted predictor has MCC = -1.

3.1.2. Results

We compared our method against other popular sparse coding methods including original Sc, ScMP, LapSc, and GraphSc. The existing somatic mutation identification method using the original features proposed by Ding et al. in [13] was also compared. The boxplots of recall, precision, accuracy, F-score and MCC of 10-fold cross-validation on the training data set are shown in Fig. 1(a–e), respectively. We observed that in all the performance measures, DisScMM outperformed the other methods significantly in terms of both the median value and the Q value. The significantly higher recall and precision of DisScMM over the other methods suggest that the improved performance is not the result of a better tradeoff, but the result of an overall better method. We also observed that the unsupervised single general graph-based sparse coding methods, i.e. GraphSc, LapSc and ScMP, had comparable performance. From these boxplots, it is not very surprising to see that the original Sc provided almost the poorest performance since the Sc function ignored the connections among the data samples. Moreover, the results of the method proposed by Ding et al. [13] without any coding procedure is slightly inferior to Sc.

Fig. 2 summarizes the recall, precision, accuracy, F-score and MMC for the proposed DisScMM and the other methods on the test dataset. According to Fig. 2, there is a significant difference between the recall and precision scores for all the methods, which

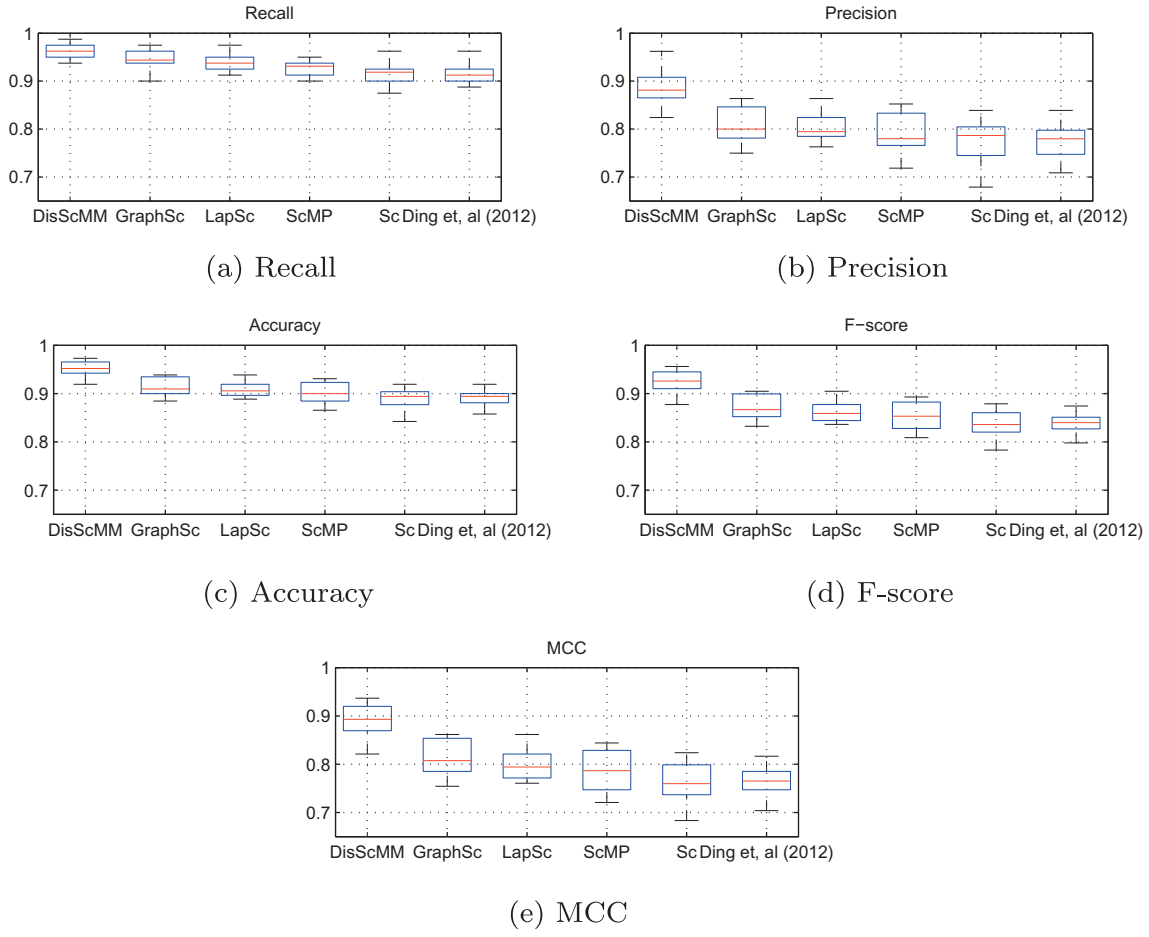

Fig. 1. Boxplots of recall, precision, accuracy, F-score and MMC of the training set 10-fold cross-validation on the somatic mutation identification.

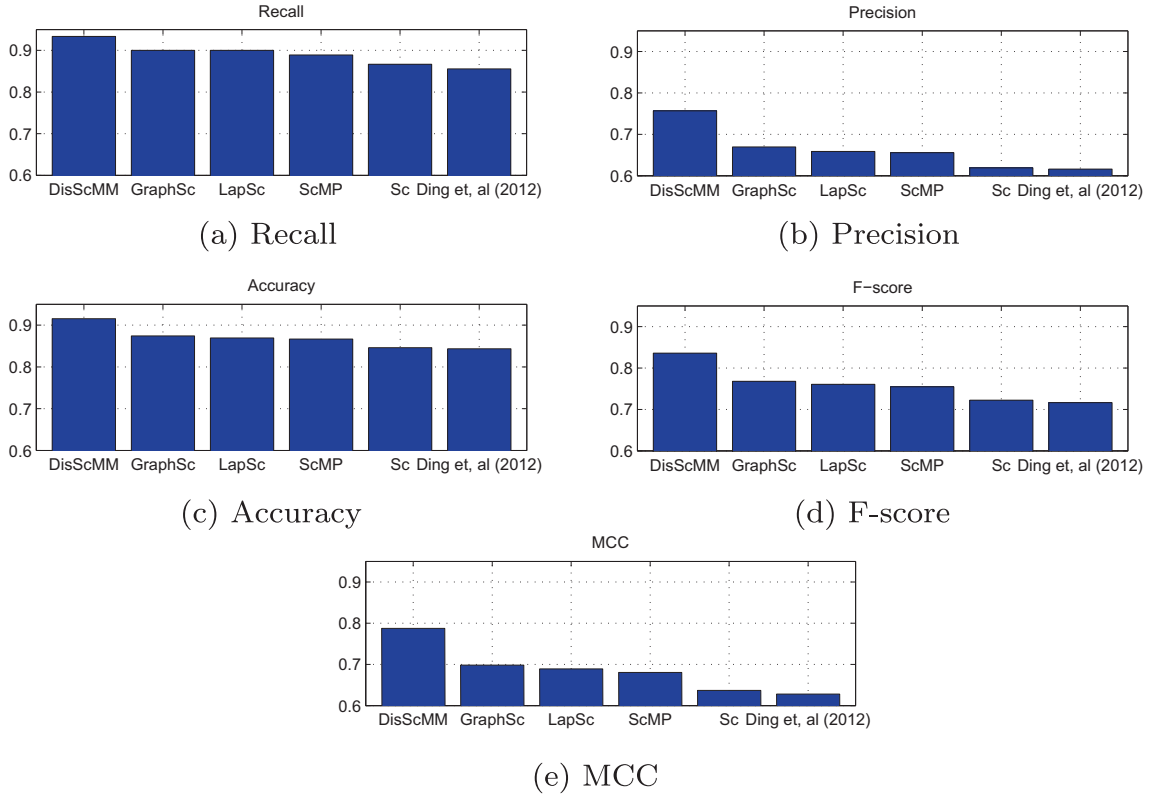

Fig. 2. The recall, precision, accuracy, F-score and MMC scores on the somatic mutation identification test set.

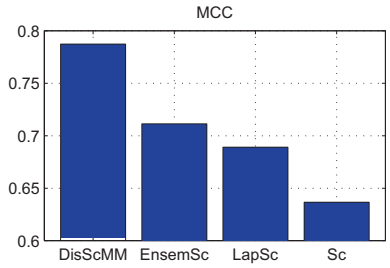

Fig. 3. Comparison to ensemble sparse coding method.

is consistent with the observations reported in the 10-fold cross-validation of the training dataset. The possible reason is the significantly unbalanced number of positive and negative samples. Second, we observe that for all the measurements, DisScMM outperforms GraphSc, LapSc, ScMP and Sc significantly. Fig. 2 also shows that the sparse coding methods with manifold regularization, such as GraphSc, LapSc and ScMP, outperform sparse coding methods without manifold regularization. Our DisScMM-based somatic mutation identifying method achieves the best somatic mutation identifying performance, which demonstrates the effectiveness of DisScMM for this task. Moreover, GraphSc, LapSc and ScMP achieve much better performance than the original Sc, which shows the usefulness of regularizing the sparse code with the nearest graphs. Moreover, the single manifold-based sparse coding methods, GraphSc, LapSc, ScMP, and method using the original feature without any coding perform similarly.

It is our general conclusion that ensemble classifiers outperform single classifiers. We further compared our method with the Ensembling Sparse Coding (EnsemSc) method that works by concatenating the sparse codes learned by Sc and LapSc. The experimental results are presented in Fig. 3. The figure shows that EnsemSc can significantly improve on the performance of the original sparse coding methods but it still does not outperform DisScMM. A possible reason for this is that both Sc and LapSc are unsupervised methods, ignoring class information that is provided by the class labels. Thus, even when Sc and LapSc are combined, the class information is still missing, while DisScMM utilizes the class information effectively in its sparse code learning procedure.

3.2. Experiment II: Breast tumor classification from ultrasound images

Medical examination based on ultrasound imaging is indispensable for the early detection and treatment of breast cancers [14,30–33]. Thus, developing an automated differential diagnosis system that classifies a given breast tumor as benign or malignant could play an important role in cancer detection. In this experiment, we evaluate the performance of our algorithm on a task of breast tumor classification task from ultrasound images.

3.2.1. Dataset and setup

We collected 340 ultrasound images to evaluate the proposed tumor classification methods. Each ultrasound image included a biopsy-proven tumor (a carcinoma, a fibroadenoma, or a cyst), where the carcinoma was a malignant tumor and the fibroadenoma and cyst were benign tumors [34,35]. The tumor border is delineated manually. The dataset contains 220 carcinomas, 60 fibroadenomas, and 60 cysts. Given an ultrasound image, we extracted 208 features and presented them in a feature vector, x. The 208 features consisted of the K-α related and conventional features [14]. 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 dataset was firstly divided randomly into five subsets and then four subsets were used for training, while the remaining subset was used for testing. We repeated the cross-validation procedure five times.

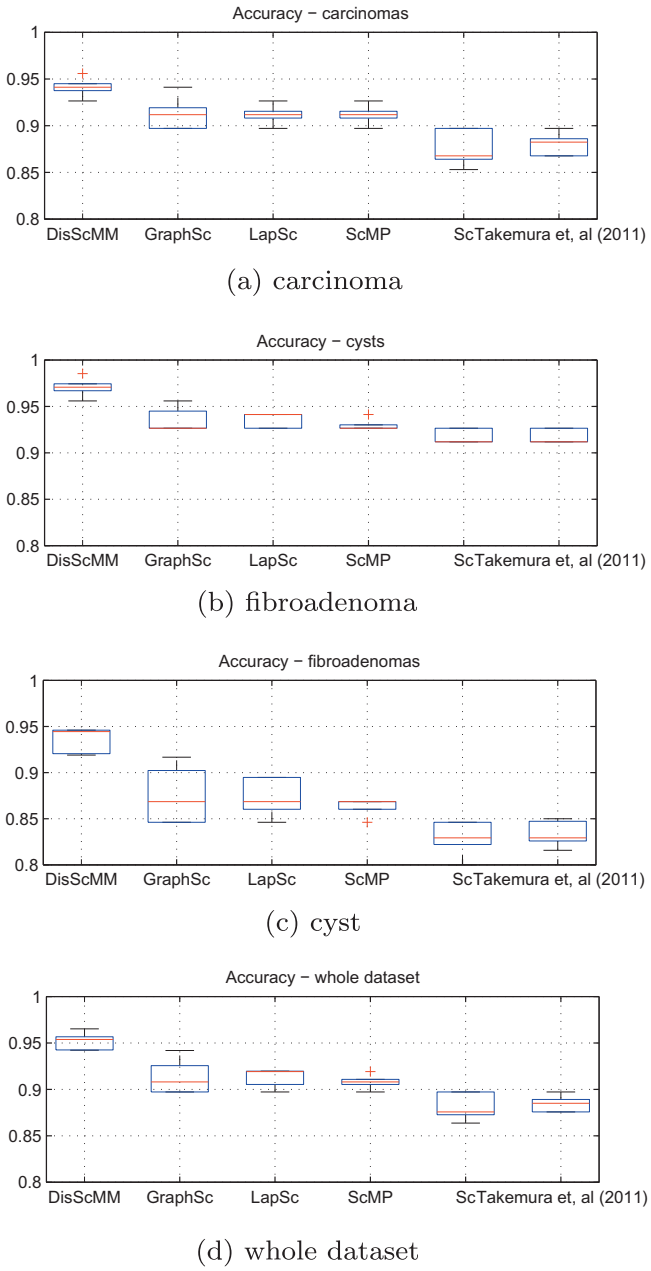

Fig. 4. Boxplots of the accuracies of the 5-fold cross-validation on different tumors on the ultrasonic breast tumor image set.

3.2.2. Results

Besides the other sparse coding methods, we also compared our method against the existing method using the original features proposed by Takemura et, al (2011) [14]. Fig. 4 shows the boxplots of the classification accuracies obtained by different methods on the ultrasound breast tumor image dataset. As shown in Fig. 4, our method achieved much better results than the state-of-the-art sparse coding methods and the original features. There are two possible reasons to explain why our method is superior to the other methods:

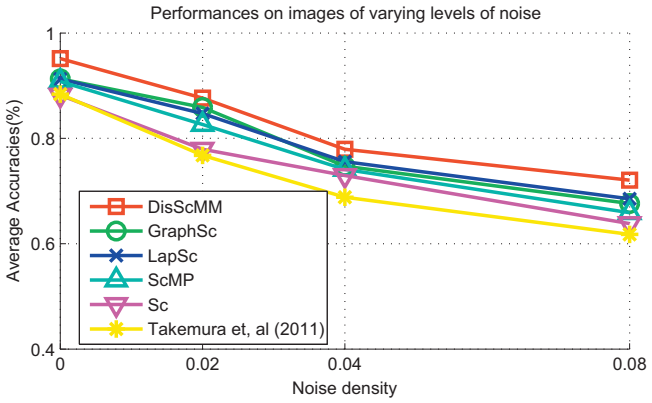

Fig. 5. Classification performance on images with varying levels of noise.

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

To show the performance of the proposed method when there is noisy data, additional comparative experiments were performed on ultrasound breast tumor images with varying levels of noise. We added noise of salt and pepper types to the images with different noise densities (0.02, 0.04, and 0.08). Average accuracies on images with varying levels of noise using sparse coding methods are given in Fig. 5. The figure shows that the performances of DisScMM decreased when higher level of noise was added to the images. From this figure, we can get the conclusion that DisScMM performs worse in the noisy data, but it still outperforms baseline methods in the noisy environment.

4. Conclusion and future work

In this paper, we proposed a novel discriminative sparse coding method to address the data representation and classification problem. Multiple manifolds are constructed for different classes. Class-conditioned sparse coding is 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. Please note that in this paper, we discuss the supervised sparse coding problem, which assumes that all the training samples are labeled. Recently, the semi-supervised learning problem has been proposed to handle the training set with both labeled and unlabeled samples [36–38]. In the future, we will try to extend our method to the semi-supervised scene.

Acknowledgments

This work was supported by grants from National Key Laboratory for Novel Software Technology, Nanjing University (Grant No. KFKT2012B17), 2011 Qatar Annual Research Forum Award (Grant No. ARF2011), and King Abdullah University of Science and Technology (KAUST), Saudi Arabia.

References

[1] H. Lee, A. Battle, R. Raina, A.Y. Ng, Efficient sparse coding algorithms, in: NIPS, NIPS, 2007, pp. 801–808.
[2] J. Eggert, E. Köner, Sparse coding and nmf, in: IEEE International Conference on Neural Networks - Conference Proceedings, vol. 4, 2004, pp. 2529–2533.
[3] I. Ohiorhenuan, F. Mechler, K. Purpura, A. Schmid, Q. Hu, J. Victor, Sparse coding and high-order correlations in fine-scale cortical networks, Nature 466 (7306) (2010) 617–621.
[4] J. Mairal, F. Bach, J. Ponce, G. Sapiro, Online learning for matrix factorization and sparse coding, Journal of Machine Learning Research 11 (2010) 19–60.
[5] J. Sun, Q. Zhuo, C. Ma, W. Wang, Sparse image coding with clustering property and its application to face recognition, Pattern Recognition 34 (2001) 1883–1884.
[6] 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, Signal Process. Soc., 2010, pp. 4346–4349.
[7] 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.
[8] K. Labusch, E. Barth, T. Martinetz, Sparse coding neural gas: learning of overcomplete data representations, Neurocomputing 72 (7–9) (2009) 1547–1555.
[9] J. Murray, K. Kreutz-Delgado, Sparse image coding using learned overcomplete dictionaries, in: Machine Learning for Signal Processing XIV – Proceedings of the 2004 IEEE Signal Processing Society Workshop, 2004, pp. 579–588.
[10] S. Gao, I. Tsang, L.-T. Chia, P. Zhao, Local features are not lonely—Laplacian sparse coding for image classification, in: 2010 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2010, pp. 3555 –3561.
[11] S. Gao, I.-H. Tsang, L.-T. Chia, Laplacian sparse coding, hypergraph laplacian sparse coding, and applications, IEEE Transactions on Pattern Analysis and Machine Intelligence 35 (1) (2013) 92–104.
[12] K.N. Ramamurthy, J.J. Thiagarajan, A. Spanias, Improved sparse coding using manifold projections, in: 2011 18Th IEEE International Conference on Image Processing (ICIP), IEEE International Conference on Image Processing ICIP, 2011, pp. 1237–1240.
[13] J. Ding, A. Bashashati, A. Roth, A. Oloumi, K. Tse, T. Zeng, G. Haffari, 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.
[14] 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.
[15] J. Lu, Y. Tan, G. Wang, Discriminative multi-manifold analysis for face recognition from a single training sample per person, IEEE Transactions on Pattern Analysis and Machine Intelligence doi:10.1109/TPAMI.2012.70.
[16] J. Lu, Y.-P. Tan, G. Wang, Discriminative multi-manifold analysis for face recognition from a single training sample per person, in: Proceedings of the IEEE International Conference on Computer Vision, 2011, pp. 1943–1950.
[17] H. Sunwoo, Multi-degree reduction of béier curves for fixed endpoints using lagrange multipliers, Computational and Applied Mathematics 32 (2) (2013) 331–341.
[18] S. Drury, S. Loisel, Sharp condition number estimates for the symmetric 2-lagrange multiplier method, Lecture Notes in Computational Science and Engineering 91 (2013) 255–261.
[19] J. Kwak, H. Cho, S. Shin, O. Bauchau, Improved finite element domain decomposition method using local and mixed lagrange multipliers, in: 54th AIAA/ASME/ASCE/AHS/ASC Structures, Structural Dynamics, and Materials Conference, 2013.
[20] S. Brogniez, C. Farhat, E. Hachem, A high-order discontinuous Galerkin method with lagrange multipliers for advection-diffusion problems, Computer Methods in Applied Mechanics and Engineering 264 (2013) 49–66.
[21] F. Tompkins, P. Wolfe, Image analysis with regularized laplacian eigenmaps, in: Proceedings – International Conference on Image Processing, ICIP, 2010, pp. 1913–1916.
[22] M. Dostanic, Regularized trace of the inverse of the Dirichlet Laplacian, Communications on Pure and Applied Mathematics 64 (8) (2011) 1148–1164.
[23] C. Wang, X. He, J. Bu, Z. Chen, C. Chen, Z. Guan, Image representation using Laplacian regularized nonnegative tensor factorization, Pattern Recognition 44 (10-11) (2011) 2516–2526.
[24] P. Perry, M. Mahoney, Regularized laplacian estimation and fast eigenvector approximation, in: Advances in Neural Information Processing Systems 24: 25th Annual Conference on Neural Information Processing Systems 2011, NIPS 2011, 2011.
[25] C. McCourt, D. McArt, K. Mills, M. Catherwood, P. Maxwell, D. Waugh, P. Hamilton, J. O'Sullivan, M. Salto-Tellez, Validation of next generation sequencing technologies in comparison to current diagnostic gold standards for BRAF, EGFR and KRAS mutational analysis, PLoS ONE 8(7).
[26] J. Rinke, V. Schäer, M. Schmidt, J. Ziermann, A. Kohlmann, A. Hochhaus, T. Ernst, Genotyping of 25 leukemia-associated genes in a single work flow by next-generation sequencing technology with low amounts of input template dna, Clinical Chemistry 59 (8) (2013) 1238–1250.
[27] G. Gelderman, L. Contreras, Discovery of posttranscriptional regulatory rnas using next generation sequencing technologies, Methods in Molecular Biology (Clifton, NJ) 985 (2013) 269–295.
[28] S. Yost, H. Alakus, H. Matsui, R. Schwab, K. Jepsen, K. Frazer, O. Harismendy, Mutascope: sensitive detection of somatic mutations from deep amplicon sequencing, Bioinformatics 29 (15) (2013) 1908–1909.

[29] 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.
[30] R. English, J. Li, A. Parker, D. Roskell, R. Adams, V. Parulekar, J. Baldwin, Y. Chi, J. Noble, A pilot study to evaluate assisted freehand ultrasound elasticity imaging in the sizing of early breast cancer: A comparison of b-mode and AFUSON elasticity ultrasound with histopathology measurements, British Journal of Radiology 84 (1007) (2011) 1011–1019.
[31] S. Mohammed, G. Meloni, M. Pinna Parpaglia, V. Marras, G. Burrai, F. Meloni, S. Pirino, E. Antuofermo, Mammography and ultrasound imaging of preinvasive and invasive canine spontaneous mammary cancer and their similarities to human breast cancer, Cancer Prevention Research 4 (11) (2011) 1790–1798.
[32] K. Hoyt, A. Sorace, R. Saini, Volumetric contrast-enhanced ultrasound imaging to assess early response to apoptosis-inducing anti-death receptor 5 antibody therapy in a breast cancer animal model, Journal of Ultrasound in Medicine 31 (11) (2012) 1759–1766.
[33] S. Bachawal, K. Jensen, A. Lutz, S. Gambhir, F. Tranquart, L. Tian, J. Willmann, Earlier detection of breast cancer with ultrasound molecular imaging in a transgenic mouse model, Cancer Research 73 (6) (2013) 1689–1698.
[34] D. Valent, J. Augsburger, Z. Correa, Embyronal carcinoma of testis metastatic to ciliary body presenting as spontaneous hyphema and painful secondary glaucoma, Retinal Cases and Brief Reports 7 (1) (2013) 105–107.
[35] M. Singh, D. Nath, K. Agarwal, Primary hydatid cyst of the breast masquerading as a fibroadenoma: a case report, Journal of Clinical and Diagnostic Research 6 (5) (2012) 886–887.
[36] M. Keyvanpour, M. Imani, Semi-supervised text categorization: exploiting unlabeled data using ensemble learning algorithms, Intelligent Data Analysis 17 (3) (2013) 367–385.
[37] M. Frasca, A. Bertoni, M. Re, G. Valentini, A neural network algorithm for semi-supervised node label learning from unbalanced data, Neural Networks 43 (2013) 84–98.
[38] T. Cupertino, L. Zhao, Semi-supervised learning using random walk limiting probabilities, Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics) 7952 LNCS (PART 2) (2013) 395–404.