Engineering Applications of Artificial Intelligence 25 (2012) 1062–1072

Contents lists available at SciVerse ScienceDirect

Engineering Applications of Artificial Intelligence

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

# Twin support vector machines and subspace learning methods for microcalcification clusters detection

Xinsheng Zhang [a,*], Xinbo Gao [b]

[a] Xi'an University of Architecture and Technology, No. 13, YanTa Road, Xi'an, Shaan xi 710055, China
[b] Xidian University, No.2, South Taibai Road, Xi'an, Shaanxi 710071, China

ARTICLE INFO

*Article history:*
Received 19 April 2011
Received in revised form
18 October 2011
Accepted 3 April 2012
Available online 5 May 2012

*Keywords:*
Microcalcification
Subspace learning
Twin support vector machines
Tensor analysis

ABSTRACT

This paper presents a novel framework for microcalcification clusters (MCs) detection in mammograms. The proposed framework has three main parts: (1) first, MCs are enhanced by using a simple-but-effective artifact removal filter and a well-designed high-pass filter; (2) thereafter, subspace learning algorithms can be embedded into this framework for subspace (feature) selection of each image block to be handled; and (3) finally, in the resulted subspaces, the MCs detection procedure is formulated as a supervised learning and classification problem, and in this work, the twin support vector machine (TWSVM) is developed in decision-making of MCs detection. A large number of experiments are carried out to evaluate and compare the MCs detection approaches, and the effectiveness of the proposed framework is well demonstrated.



## 1. Introduction

Breast cancer is probably the most common non-skin cancer and one of the major causes of death in women as well as a number of men. Aiming at breast cancer detection at the early stage, the digital mammography has become one of the most effective methods. In digital mammography data, an important sign of breast cancer is the existence of microcalcification clusters (MCs), which appear in 30–50% of mammo- graphically diagnosed cases with tiny bright spots of a different morphology. Microcalcifications are small calcifications of different shapes and densities, approximately 0.1–1 mm in diameter. Isolated microcalcifications are not dangerous, but a microcalcification cluster, which is a region containing three or more microcalcifications per $5\,\text{mm} \times 5\,\text{mm}$ area, might be an early sign of breast cancer (Kopans, 2006).

Because of its importance in breast cancer diagnosis, accurate detection of MCs has become a key research and application task, and a number of approaches have recently been developed, which have been greatly assisting doctors and radiologists in diagnosing breast cancer (Cheng et al., 2003). Among them, apart from focusing on image segmentation and specification of regions of interest (ROI), several methods have been proposed, such as classical image filtering and local thresholding (Bagci and Cetin, 2002; Nakayama et al., 2006), and techniques based on mathematical morphology (Halkiotis et al., 2007; Mossi and Albiol, 1999), fractal models (Bocchi et al., 2004), optimal filters (Gulsrud and Husoy, 2001), wavelet analysis and multiscale analysis (Bagci and Cetin, 2002; Rezai-rad and Jamarani, 2005; Song et al., 2006). Various classification approaches have also been presented to characterize MCs, such as rule-based systems (Riyahi-Alam et al., 2004), fuzzy logic systems (Cheng et al., 1998; Cheng et al., 1998; Cheng et al., 2004; Jiang et al., 2005), statistical methods based on Markov random fields (D'Elia et al., 2004; Lee and Chen, 1996), and support vector machines (SVMs) (D'Elia et al., 2004; El-Naqa et al., 2002). In the past decade, most work reported in the literature has employed neural networks in MCs characterization (Bocchi et al., 2004; Halkiotis et al., 2007; Hernandez-Cisneros and Terashima-Marin, 2006; Papadopoulos et al., 2005; Rezai-rad and Jamarani, 2005; Yu and Guan, 2000). With the development of SVMs, various SVMs have been designed to categorize ROIs (El-Naqa et al., 2002). SVMs are promising as it concerns on the trade-off between statistical structural complexity and empirical risk, though it also encounters some problems. One of the most popular explanations of SVM classifiers is probably the maximum margin that attempts to reduce generalization error by maximizing the margin between two disjoint half planes. The resulting optimization task involves the minimization of a convex quadratic objective function subject to linear inequality constraints. Recently, Jayadeva et al. (2007) have proposed a nonparallel plane classifier for binary data classification, termed as twin support vector machine (TWSVM). This classifier aims to generate two nonparallel planes such that each plane is closer to one of the two classes and is as far as possible from the other. TWSVM solves two quadratic programming problems (QPPs) of smaller size

* Tel.: +86-13152031863.
*E-mail addresses:* xinsheng.zh@hotmail.com (X. Zhang), xbgao.xidian@gmail.com (X. Gao).

0952-1976/$ - see front matter © 2012 Elsevier Ltd. All rights reserved.
http://dx.doi.org/10.1016/j.engappai.2012.04.003

instead of large sized ones as we do in traditional SVMs. This makes TWSVM work almost four times faster than the standard SVM classifier with the linear separable problem.

Although computer-aided mammography has been studied over two decades, automated detection of MCs remains very difficult. This is because of the existences of film emulsion error, digitization artifacts, or anatomical structures such as fibrous strands, breast borders, or hypertrophied lobules, which look like microcalcifications in the breast normal tissues, which often cause high false positives. Meanwhile, in some dense tissues, and/or skin thickening, especially in the breasts of younger women, suspicious areas are almost invisible. The dense tissues, especially in younger women may easily be regarded as microcalcification, which often cause to a false positive. That is the major problem with most of the detection algorithms.

To improve the performance of existing MCs detection algorithms, we propose a novel framework to detect MCs. In the framework, we employ TWSVM as the classifier to distinguish MCs from other ROIs in images. The classifier is first trained by the manually labeled image blocks of mammograms from the digital database for screening mammography (DDSM) (Rose et al., 2006). Then it is used to detect other ROIs. The effectiveness of TWSVM is then evaluated.

In the experiment, a set of 267 images of clinical mammograms from the widely recognized DDSM database is selected to extract ROIs as the *test bed*, in which 2231 MCs exist. These mammograms were divided into two subsets: the first set is for training and validation in training stage and the other for test in the testing stage. Compared with several popular existing methods, this newly proposed approach achieves better performance. In evaluation stage, curves of receiver operating characteristic (ROC) are used. The best average sensitivity is as high as $94.41 \pm 1.16\%$ with respect to $8.32 \pm 1.04\%$ false positive rate and the area under the ROC curve, $\mathrm{Az}{=}0.9677 \pm 0.0574$ in each testing phase.

In our paper, the main contribution is to propose an application framework for MCs detection by incorporating the subspace learning algorithms and TWSVM. Although it is commonly used in pattern recognition, to the best of our knowledge, there are few works employing such a framework to detect MCs in mammograms. In our research, we apply subspace learning to extract discriminant features rather than semantic features. Then TWSVM and SVM-based classifier are trained to distinguish the blocks with MCs from others. By comparing some available subspace learning methods, such as PCA, LAD, TSA and GTDA, and classifiers, such as SVM and TWSVM, some meaningful conclusions are reached, which provide guidance for the proposed framework of MCs detection in mammograms.

The rest of this paper is organized as follows: background of subspace learning and TWSVM is given in Sections 2 and 3, respectively. Thereafter, the novel TWSVM based MCs detection algorithm is formulated in Section 4. The new approach is then evaluated as described in Section 5, and accordingly, experimental results are reported in Section 6. Finally, conclusions are drawn in Section 7.

## 2. Subspace learning

Traditionally, subspace learning approaches (Fu et al., 2008), such as principal component analysis (PCA) and linear discriminant analysis (LDA), scan any type of input data as vectors, and then transform the high dimensional data into a low-dimensional subspace. However, in the real world, an image is intrinsically a matrix, or a 2-order tensor. The extracted feature of an object often has some specialized spatial structures, and such structures are in the form of second-order or even high-order tensors. And also, because of the under sample problems (USP), i.e., the dimensionality of input feature space is much higher than the number of training samples (in our task the training sample number $m$ is much less than the input sample feature dimensionality, $m \ll 115 \times 115$), PCA and LDA do not always work as well as expected. To overcome this problem, multilinear algebra, the algebra of higher-order tensors, has recently been applied to analyze multifactor structures of image ensembles. Furthermore, a set of promising approaches have been developed, such as tensor space analysis (TSA) (He et al., 2005) and general tensor discriminant analysis (GTDA) (Tao et al., 2007). TSA treats an image as a second-order tensor in $R^{n_1} \otimes R^{n_2}$, where $R^{n_1}$ and $R^{n_2}$ are the two vector spaces. The relationship between row vectors can be naturally characterized by TSA and GTDA, and they can detected by learning the intrinsic local geometrical structure information in the tensor space. This structure information of objects in pattern recognition research is a reasonable constraint to reduce the number of unknown parameters used to represent a learning model.

In this paper, the symbols in bold lowercase represent vectors, such as $\boldsymbol{x}, \boldsymbol{y}$; the bold uppercase symbols represent matrices, such as $\boldsymbol{A}, \boldsymbol{B}, \boldsymbol{C}$; the *italic* uppercase symbols represent tensor objects, such as $\mathcal{X}$; and the *italic* lowercase symbols stand for scale numbers, such as $y$, $c$.

### 2.1. PCA and LDA

PCA is a typical linear dimensionality reduction algorithm. The basic idea of PCA is to project the data along the directions of maximal variances so that the reconstruction error can be minimized. Given a set of data points or patterns $\boldsymbol{x}_1,\boldsymbol{x}_2,\ldots,\boldsymbol{x}_n$, let $\boldsymbol{w}$ be the transformation vector and $y_i = \boldsymbol{w}^T\boldsymbol{x}_i$. The objective function of PCA is

$$\mathbf{w}_{opt} = \arg\left\{\max_{\mathbf{w}} \sum_{i=1}^{n} (y_i - \overline{y})^2\right\} = \arg\left\{\max_{\mathbf{w}} \mathbf{w}^T\mathbf{C}\mathbf{w}\right\}, \tag{1}$$

where $\overline{y} = \frac{1}{n}\sum y_i$ and $\boldsymbol{C}$ is the data covariance matrix. The basis functions of PCA are the eigenvectors of the data covariance matrix associated with the largest eigenvalue.

Compared with PCA seeking directions that are efficient for representation, LDA seeks directions that are efficient for discrimination. If we have a set of $n$ samples $\boldsymbol{x}_1,\boldsymbol{x}_2,\ldots,\boldsymbol{x}_n$, belonging to $k$ classes. The objective function of LDA is as follows:

$$\mathbf{w}_{opt} = \arg\left\{\max_{\mathbf{w}} \frac{\mathbf{w}^T\mathbf{S}_B\mathbf{w}}{\mathbf{w}^T\mathbf{S}_W\mathbf{w}}\right\} = \arg\left\{\max_{\mathbf{w}} \frac{tr(\mathbf{w}^T\mathbf{S}_B\mathbf{w})}{tr(\mathbf{w}^T\mathbf{S}_W\mathbf{w})}\right\}, \tag{2}$$

$$\mathbf{S}_B = \frac{1}{n}\sum_{i=1}^{k} n_i(\mathbf{m}^{(i)} - \mathbf{m})(\mathbf{m}^{(i)} - \mathbf{m})^T, \tag{3}$$

$$\mathbf{S}_W = \frac{1}{n}\sum_{i=1}^{k} \left(\sum_{j=1}^{n_i} (\mathbf{x}_j^{(i)} - \mathbf{m}^{(i)})(\mathbf{x}_j^{(i)} - \mathbf{m}^{(i)})^T\right), \tag{4}$$

where $\boldsymbol{m}$ is the total sample mean vector, $n_i$ is the number of samples in the *ith* class, $\boldsymbol{m}^{(i)}$ is the average vector of the *ith* class. We call $\boldsymbol{S}_W$ the within-class scatter matrix and $\boldsymbol{S}_B$ the between-class scatter matrix.

### 2.2. Tensor space analysis

TSA is a new technique which learns a tensor subspace which respects the geometrical and discriminative structures of the original data space. Let $\mathbf{X} \in R^{n_1 \times n_2}$ denote an image of size $n_1 \times n_2$, where $\mathcal{X}$ can be thought of as a 2-order tensor in tensor space