Journal of Fiber Bioengineering and Informatics 8:1 (2015) 161–169
doi:10.3993/jfbi03201516

Hybrid Subspace Fusion for Microcalcification Clusters Detection ⋆

Xinsheng Zhang*, Hongyan He, Naining Cao, Zhengshan Luo

School of Management, Xi'an University of Architecture and Technology, Xi'an 710055, China

Abstract

Early detection of breast cancer, a significant public health problem in the world, is the key for improving breast cancer early prognosis. Mammography is considered the most reliable and widely used diagnostic technique for early detection of breast cancer. However, it is difficult for radiologists to perform both accurate and uniform evaluation for the enormous mammograms with widespread screening. Microcalcification clusters is one of the most important clue of the breast cancer, and their automated detection is very helpful for early breast cancer diagnosis. Because of the poor quality of the mammographic images and the small size of the microcalcifications, it is a very difficult task to perform detecting the early breast cancer. In this paper, we propose a novel approach based on hybrid subspace fusion for detection microcalcification clusters, and successfully apply it to detection task in digital mammograms. In such a system, subspace learning algorithms will be selectively fused according to the ability of preserving the classification information. Experimental results show that the proposed method improved the performance and stability of microcalcification cluster detection and could be adapt to the noise environments better. The proposed methods could get satisfactory results on sensitivity and reduce false positive rate, which provide some new ideas and methods for the research and development of computer-aided detection system in the breast cancer detection community.

Keywords: Subspace Learning; Data Fusion; Micalcification Cluster; Support Vector Machine; Digital Mammograms

1 Introduction

In digital mammograms, an important sign of the early breast cancer is the existence of Microcalcification Clusters (MCs). One of the key techniques for early diagnosis of the breast cancer is to detect MCs and to judge whether they are malignant or not in mammograms. However, there is only about 3% information in mammograms, which can be seen with the naked eyes. Due to the most details in mammograms cannot been perceived by human eyes, it is even very

⋆Project supported by the National Natural Science Foundation (No. 61271278) and the Key Discipline of Shaanxi Province (No. E08001).

*Corresponding author.
Email address: xinsheng.zh@gmail.com (Xinsheng Zhang).

difficult for a skillful radiologist to find the sign of early breast cancer, i.e., micalcification clusters, as a result missing the best time for treatment. To detect early sign of this disease and we need to do research on how to build a computer-aided diagnosis system to aid doctors to diagnose breast cancer in early stage. Machine learning techniques have been successfully applied to it [1–5].

Most of these methods learn hypotheses from a large amount of diagnosed mammogram samples, i.e., the data collected from a number of necessary medical examinations with the corresponding diagnoses made by the doctors, to aid the other doctors to do a diagnosis in the future. Because of the big dimensionality of the image samples, there always are an overfitting problem caused the curse of dimensionality in those machine learning methods [6].

This curse of dimensionality is compounded by extra complexity introduced by data set heterogeneity. Heterogeneous data sets usually require more sample data points than homogeneous data sets for machine learning performance. Combining heterogeneous data types for classification is a difficult machine learning problem, but one that has shown promise in bio-informatics applications. Data fusion is the study of optimal information processing in heterogeneous data environments through intelligent integration of heterogeneous data. The development of data fusion was motivated chiefly by the increasing demand for more accurate information, more practical and robust procedures to manage data efficiently, and improved system reliability and performance.

Decision fusion has the ability to capture general trends in data while remaining robust to the effects of noise. As the noise and complexity of a data set increase, it becomes easier for a classifier model to overfit the data rather than capture the underlying data trends. Currently, researchers have proposed a lot of microcalcification detection methods based on ensemble learning. As Li *et al.* [7] proposed a Co-Forest-based ensemble learning method to detect microcalcifications when labeled samples is limited. Jesneck *et al.* [8] proposed an optimized breast cancer detection method based on heterogeneous multi-source information fusion to improve diagnostic performance, which can combine heterogeneous information, such as from different modalities, feature categories, or institutions. Experimental results showed that the decision fusion directly optimized clinically significant performance measures, such as AUC and pAUC, and sometimes outperformed two well known machine-learning techniques. But a single type of image features was not always a good choice to improve the performance of detecting microcalcification clusters. To improve the performance of detection system, Massimo D. S. *et al.* [9] proposed a microcalcification cluster detection algorithm based on information fusion with a multiple expert system. In this system, there are several experts, some of which are devoted to classify the single microcalcifications and others are aimed to classify the cluster considered as a whole. Then the knowledge of multiple experts are fused together to make a final overall decision.

Methods discussed above have achieved some good results by fusing different learning methods, but most of these methods are carried out the classification task based on the certain fixed characteristics or domain knowledge, without considering how to get these features or dynamic domain knowledge. To take advantage of decision fusions strengths and to keep more information, based on results of previous research, subspace learning and ensemble learning of the latest research study, we pursued a data fusion approach that was a hybrid of the feature fusion and decision fusion techniques to perform the MCs detection task. We first extracted features from the raw data (carefully selected sample database). We applied these features to a local classifier and thresholds to these classifiers outputs, creating local binary decisions. We then combined these