Feature Selection Using General Regression Neural Networks for the Automatic Detection of Clustered Microcalcifications

Songyang Yu and Ling Guan

School of Electrical & Information Engineering
University of Sydney, NSW 2006, Australia

Abstract

General regression neural networks (GRNNs) are proposed for selecting the most discriminating features for the automatic detection of clustered microcalcifications in digital mammograms. Previously, We have designed an image processing system for detecting clustered microcalcifications. The system uses wavelet coefficients and feed forward neural networks to identify possible microcalcification pixels and a set of structure features to locate individual microcalcifications. In this work, more features are extracted, and the most discriminating features are selected through the analysis of the GRNNs. The selected features are incorporated into our image processing system and applied to a database of 40 mammograms (Nijmegen database) containing 105 clusters of microcalcifications. Free response operating characteristics (FROC) curves are used to evaluate the performance. Results show that, by incorporating the proposed feature selection scheme, the performance of our system is improved significantly.

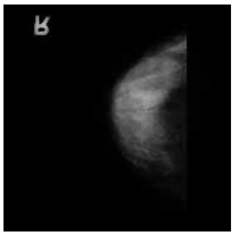
1. Introduction

Cancer of the breast is now one of the common forms of cancer diagnosed in women. In Australia, 67 out of 10,000 women were diagnosed breast cancer in 1992 [1]. Neither the cause of breast cancer nor the means of preventing the disease are well understood. At present, early detection of breast cancer is the only way to reduce the breast cancer mortality and enhance the cure rate.

One of the important early symptoms of breast cancer in the mammograms is the appearance of microcalcification clusters. They have a higher X-ray attenuation than normal breast tissue and appear as a group of small, localized granular bright spots in the mammograms. A typical mammogram with clusters of microcalcifications is shown in Figure 1(a), and the magnified version of a cluster of microcalcifications in Figure 1(b).

In our previous work, we have developed a computer aided diagnosis (CAD) system for the automatic detection of clustered microcalcifications based on wavelet coefficients and neural networks [2][3]. Our CAD system has two main steps. First, possible microcalcification pixels in the mammogram are segmented out and labelled into possible individual microcalcification objects by their spatial connectivity. This is achieved by using wavelet coefficients, gray level statistical features and back propagation neural networks. Among these possible individual microcalcification objects, there are a lot of false detections due to the noise, blood vessels and dense breast tissue in the mammogram. In order to eliminate these false detections, in the second step, individual microcalcification objects are detected based on a set of nine structure features by using feed forward neural networks. The system was applied to the database provided by the university hospital of Nijmegen [4]. By using a free response operating characteristics (FROC) curve to evaluate the performance, our CAD system achieved 75% mean true positive detection rate at the cost of 0.5 false positive per image.

In order to improve the performance of our CAD system, a critical step is to add more effective features for the detection of individual microcalcification objects. In this paper, in addition to the nine structure features already used, four shape moment features [5], seven invariant moment features [6] and ten second order histogram related features [7] are added to describe the possible individual microcalcification objects. Then a feature selection method based on the general regression neural networks (GRNNs) [8] is proposed to select the most discriminating feature sets. The selected features are then incorporated into our CAD system. Experimental results show that by using these features, the performance of our CAD system increased to 90% mean true positive detection rate at the cost of 0.5 false positive per image. In particular, our system outperforms Karssemeijer's [4], one of the best in the literature. In the process, it is discovered that features related to the second order histogram are of vital importance for the detection of clustered microcalcification.



(a)

(b)

Figure 1. (a) A mammogram from the database. (b) The magnified view of a cluster of microcalcifications.

2. Using GRNNs to Select the Most Discriminating Features

For a given pattern recognition problem, there is a large number of features which could be extracted from the objects to be classified. Therefore, it is necessary to select a finite set of features that has the most discriminating power for the classification of the objects.

Given a set of N dimension vectors Y which representing a pattern belonging to one of m classes, each dimension representing a feature describing some property of Y. The feature selection problem is to select a subset of n ($n \le N$) features, which contains more discriminatory information than any other subset of n features in Y.

The optimal subset of features can be determined by exhaustive testing all the possible combinations, which equals to $\sum_{i=1}^{N}\binom{N}{n}$. However, even for a moderate N and n, this is a large number which makes an exhaustive search infeasible. Instead of using exhaustive search, there are three other methods to deal with this problem. The first is by experience which has been used by most researchers in this field, such as [4]. The second is feature transformation, which is implemented in such a way that the transformed features have less dimension than the original features, but have more discriminating power. Principal component analysis [9] is one of the well known method of this category. The third is to organize the search method to reduce the number of feature sets to be evaluated. Sequential Forward Selection (SFS) and Sequential Backward Selection(SBS) [10] are two methods in this category. In this paper, GRNNs are used as the vehicle to realize the SFS and SBS methods.

2.1 General Regression Neural Network

The GRNNs are memory-based feed forward networks based on the estimation of probability density function. Originally known as Nadaraya-Watson regression in the statistics literature, it is rediscovered by Donald Specht [8]. Let x be a vector random variable, y be a scalar random variable, and $f(x,y)$ represent the joint probability density function of x and y. The expected value of y given X can be computed by:

$$E[y|X] = \frac{\int_{-\infty}^{\infty} y f(X,y)\,dy}{\int_{-\infty}^{\infty} f(X,y)\,dy} \tag{1}$$

In practice, the probability density function is usually unknown. So it must be estimated from sample values of X_i and Y_i. The estimator, also called kernel function, proposed by Parzen [11] is used:

$$\begin{aligned}\hat{f}(x,y) &= \frac{1}{(2\pi)^{(p+1)/2}\sigma^{(p+1)}} \\ &\bullet \exp\left[-\frac{(Y-Y_i)^2}{2\sigma^2}\right] \\ &\bullet \frac{1}{n}\sum_{i=1}^{n}\exp\left[-\frac{(X-X_i)^T(X-X_i)}{2\sigma^2}\right]\end{aligned} \tag{2}$$

where σ is the width of the estimating kernel, n is the number of samples, and p is the dimension of the input vector X. Substituting the probability estimator in (2) into equation (1) gives the desired conditional mean of y given X:

$$Y(X) = \frac{\sum_{i=1}^{n} Y^i \exp\left[-\frac{D_i^2}{2\sigma^2}\right]}{\sum_{i=1}^{n}\exp\left[-\frac{D_i^2}{2\sigma^2}\right]} \tag{3}$$

Where D_i^2 is defined as:

$$D_i^2 = (X-X_i)^T(X-X_i) \tag{4}$$

The topology of a GRNN consists of 4 layers: the input layer, the hidden layer, the summation layer and the output layer. The function of the input layer is simply to pass the input vector variables X to all the units in the hidden layer. The hidden layer consists of all the training sample $X_1 \ldots X_n$. When an unknown pattern X is presented, the squared distance D_i^2 between the unknown pattern and the training sample is calculated and passed through the kernel function. The summation layer has two units A and B, The unit A computes the summation of $\exp\left[-D_i^2 / (2\sigma^2)\right]$ multiplied by the Y_i associated with X_i. The B unit simply computes the summation $\exp\left[-D_i^2 / (2\sigma^2)\right]$. The output unit divides A by B to provide the predication result.

2.2 Feature Extraction and Selection

Among the possible individual microcalcification objects segmented out in the first step of our CAD system, there are a lot of microcalcification like objects because of noise, blood vessels and dense breast tissue. This makes the false detection rate relatively high. In order to decrease the false detection rate, a second step processing based on a set of features of the possible microcalcification objects is conducted.

In extracting the features, a square neighbourhood of 10 pixels larger than the possible microcalcification object in diameter is used to define the background window of each object. We selected a pool of 31 features as candidates for the feature selection algorithm. The features are listed in Table 1. The 31 features can be divided into two groups: structure features and second order histogram related features. In order to calculate the second order histogram features, the second order histogram of each possible individual microcalcification objects is calculated first.

For selecting features, an error function measuring the discriminating power of the feature set being evaluated should be given first. For a pattern recognition problem, the error function is usually the mean squared classification error which is defined as:

$$E= \frac{1}{k} \sum_{i=1}^{k} \left(d_i - y_i \right)^2 \tag{5}$$

where i is the i th test data, y_i is the output of the GRNN and d_i is the desired output.

In our work, SFS and SBS methods are carried out by the GRNNs. SFS is a simple bottom up search method, starting with one feature from all the features, which gives the smallest mean squared classification error. Then a new feature from the remaining features is added at a time to the current feature set. The new feature added is the one which gives the smallest mean squared classification error compare to adding others. On the other hands, SBS is the top down counterpart of the SFS method. At the beginning, all features are included. We then discard one feature at a time. The feature discarded is the one that gives the smallest mean squared error function by removing it. Thus the feature discarded has the least discriminating power. To find the most discriminating feature set, the algorithm will stop at a point where the mean squared classification error begins to increase if further adding or discarding a feature is performed. For finding the order of the discriminating power of all features, the algorithm will continue until all the features are added or discarded.

3. Experimental Results

In order to select the most discriminating features, a training/test set of true and false individual microcalcification objects are produced from the database. The training data set consists of 174 true individual micocalcification objects and 164 false individual microcalcification objects. All the 31 features of each true or false individual microcalcification are first calculated. As a preprocessing step, these features are normalized between -1 and 1 before fed into the GRNN.

In order to use the GRNN as the classifier, we first have to choose the width of the probability σ. Since we have a limited number of training/test data. A cross validation method called leave-one-out is used. For a particular value of σ with a training/test data set of n samples, the leave-one-out method moves one sample at a time and constructs the GRNN using the remaining (n-1) samples. Then the GRNN is used to classify the removed sample. This is repeated n times, and each classification result is stored. Then the mean squared classification error of this σ is calculated. The σ which minimizes the mean squared classification error is chosen for later study. For our training/test data, the σ that gives the smallest mean squared classification error is 0.6.

After chosen σ, feature selection methods based on SFS and SBS algorithm are use to select the most discriminating features. For each feature set evaluated, a GRNN is constructed to generate the actual classification results, the leave-one-out method is used to calculate the mean square classification error for that feature set. The first 15 most discriminating features selected by using SBS and SFS algorithm respectively and the 12 features selected by both SBS and SFS are listed in Table 2. It shows that 7 out of the 10 second order histogram related features are among the 15 most discriminating features. This demonstrates that the features related to the second order histogram play an important role in the detection of individual microcalcification objects.

After the most discriminating feature sets are selected, three feed forward neural networks are trained using selected feature sets by SBS, SFS and SBS/SFS respectively. For comparison, another feed forward neural network using all the 31 features is also trained. The feed forward neural networks are then incorporated in to our CAD system to detect individual microcalcification objects. To evaluate the performance of these feature sets, FROC curve is used. The performance of the three selected feature sets comparing with using all 31 features is shown in Figure 2. The two feature sets selected by SBS and SFS performs better than all 31 features. The features selected by the SBS has the best performance. It achieves a 90 percent mean true positive rate at the cost of 0.5 false positive per image which is much better than our previous result (75% mean true positive rate at the cost of 0.5 false positive per image [3]). Shown in Figure 3 is the comparison of our results using SBS selected features with Karssemeijer's result [4] using MRF model and IPA scaling with local image feature, namely, local contrast(lc), smoothed local contrast(lcs), and the line/edge feature(lin). It clearly shows that our method with the features selected by SBS outperforms Karssemeijer's.

4. Conclusions

This paper demonstrates a method of selecting the most discriminating features for the identification of individual microcalcifications objects using the GRNNs. The results show that features related to the second order histogram play an important role in the detection of individual microcalcifications objects. By using the features selected by the GRNNs, our CAD system shows superior performance over the other reported works in the literature.

5. Acknowledgement

This work is partially supported by a grant from Australian Research Council.

6. References

[1] Breast Cancer Screening and Treatment in Australia, Report of the Senate Standing Committee on Community Affairs, June, 1994.

[2] Songyang Yu, Ling guan, Stephen Brown, Automatic Dete-

tion of Clustered Microcalcifications in Digital Mammograms Based on Wavelet Features and neural Network Classification, SPIE Medical Imaging, San Diego, USA. 1998.
[3] Songyang Yu, Ling Guan, Stephen Brown, Automatic Detection of Clustered Microcalcifications in Digitized Mammogram Films, submitted to Journal of Electronic Imaging.
[4] N. Karssemeijer, Adaptive Noise Equalization and Recognition of Microcalcifications in Mammograms, International Journal of Pattern Recognition and Artificial Intelligence Vol. 7, No. 6, pp. 1357-1376, 1993.
[5] L. Shen, R. M. Rangayyan, J.E. Leo Desautels, Shape Analysis of Mammographic Calcifications, Proceedings of the Fifth Annual IEEE Symposium on Computer-based Medical Systems, pp. 123-128, June 1992.
[6] R. C. Gonzalez, R. E. Woods, Digital Image Processing, Addison-Wesley, 1993.
[7]] R. M. Haralic, K. Shanmugam, I. Dinstien, Texture Features for Image Classification, IEEE Transactions on System, Man, and Cybernetics, Vol. SMC-3, No. 6, pp. 610-621, Nov. 1973.
[8] D. F. Specht, A General Regression Neural Network, IEEE transactions on Neural Networks, Vol. 2, No. 6, pp. 568-576, Nov. 1991.
[9] S. Haykin, Neural Networks: a comprehensive foundation, Maxwell Macmillan International, 1994.
[10] J. Kittler, Feature Set Search Algorithms, Pattern Recognition and Signal Processing, Sijthoff & Noordhoff 1978, The Netherlands.
[11] E. Parzen, On Estimation of a probability density function and mode, Ann. Math. Statist., Vol. 33, pp. 1065-1076, 1962.

Table 1: List of features in the feature selection.

Feature	Name	Description
1	Mean	average gray level, standard deviation of gray level, edge strength and background gray level of each objects.
2	Standard Deviation	
3	Edge	
4	Background(bac)	
5	Foreground Background Ratio(fbr)	fbr=mean/bac
6	Foreground Background Diffeence(fbd)	fbd=mean-bac
7	Difference Ratio(dr)	dr=(mean-bac)/(mean+bac)
8	Area	size of the object
9	Compactness	C=perimeter^2/area
10	Elongation	E=max. axis/min. axis
11-14	Shape Moment 1-4	4 Shape moment feature[5]
14-21	Invariant Moment 1-7	7 Invariant moment[6]
22	Contrast	Second order histogram related features[7]
23	Entropy	
24	Angular Second Moment	
25	Inverse Different Moment	
26	Correlation	
27	Variance	
28	Sum Average	
29	Sum Entropy	
30	Sum Variance	
31	Difference Entropy	

Table2: Features selected by SFS and SBS method.

Feature Selection Method	Selected Features
SFS	1,2,4,7,8,9,11,14,22,23,24,25,26,27,29
SBS	2,6,7,8,9,10,14,19,22,23,24,25,26,27,29
SFS & SBS	2,7,8,9,14,22,23,24,25,26,27,29

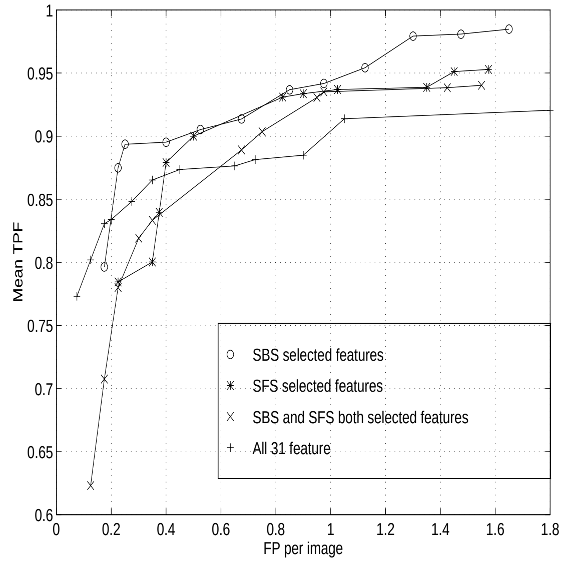


Figure 2. Comparison of FROC performance the among the feature selection methods.

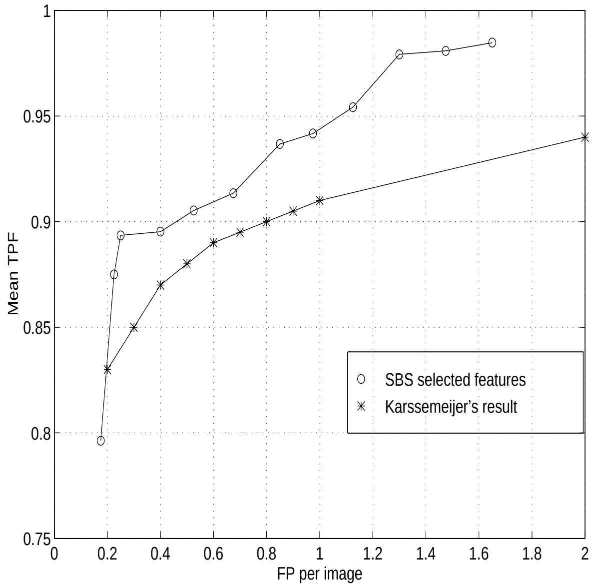


Figure 3. Comparison of FROC performance of the SBS selected features to Karssemeijer's result.