Classification of mammographic features using RBF-SA

Rafael do Espírito Santo
Pesquisador Colaborador – EP-USP;
Professor do curso de Ciência da Computação – Uninove.
São Paulo – SP [Brasil]
rafaelid@uol.com.br

Roseli de Deus Lopes
Professor Adjunto do Departamento de Engenharia Elétrica da Escola Politécnica da Universidade de São Paulo – USP.
São Paulo – SP [Brasil]
roseli@lsi.usp.br

Rangaraj M. Rangayyan
Rangaraj M. Rangayyan is a Professor with the Department of Electrical and Computer Engineering, and an Adjunct Professor of Surgery and Radiology, at the University of Calgary, Calgary, Alberta. Canada;
Calgary [Canadá]

We present in this work a new type of classes discriminator based upon nonlinear and combinational optimization techniques: radial basis functions-simulated annealing (RBF-SA). The combinational optimization method is used here as a pre-estimation of some parameters of the network classifier. We compare the classifier performance with and without pre-estimation. For training the classifiers, adopting the leave-one-out procedure, we have used case examples such as mammographic masses (malignant and benign). The classifier is trained with shape factors and edge-sharpness measures extracted from 57 regions of interest (ROI) (37 malignant and 20 benign), manually delineated, that describe mammographic masses and tumor features in terms of polygonal models for shape factors (compactness [CC], Fourier description [FF], fractional concavity [F_{CC}] and speculated index [SI]) and edge sharpness-acutance (A) . The classifier performance is compared in terms of the area under the receive operating characteristic (ROC) curve – (A). Higher values of A correspond to a better performance of classifier. Experiments with mammographic tumor and masses show that the best result of 0.9776 is obtained with RBF-SA when RBF parameters such as centers and spread matrix are pre-estimated, which is significantly better than the results obtained with no pre-estimation or only pre-estimation of the RBF centers, which are, 0.7071 and 0.9552 respectively.

Key words: Mammography. Neural networks. Optimization. Performance. Simulated annealing.

1 Introduction

Radial base functions (RBF) as network classifier: the design of a classes discriminator can be viewed as solving optimization problem known in statistics as stochastic approximation. As far as this approach is concerned, the learning process is the same as finding a surface in a multidimensional space that provides the best adaptation of data used for training the classifier. On the other hand, the ability of generalization of the classifier is like using these multidimensional surfaces to interpolate the data used for testing the classifier. These equivalences are the motivation of using RBF (HAYKIN, 1999) to design neural network that separates classes. In the context of the artificial neural network, the hidden layers provide a set of functions that constitute the bases generators of a multidimensional space. These bases permit the representation of data connected to network input in the space generated on hidden layer (HAYKIN, 1999).

Broomhead and Lowe (1988) were the first researchers to explore the design of neural network using RBF. Other works related to that kind of ensue are: Powell (1985), Moody and Darken (1989), Renals (1989) and Poggio and Girosi (1990). The architecture of a neural network implemented from the RBF has three distinct layers. The first layer is the place where the necessary data for training the network is connected. The second layer is a space that has high dimension when compared to the input layer dimension. The third layer is the output of the network. It is the place where the answers of the network are collected regarding the activations made by input data. The transformation between input and hidden layer space is nonlinear, while the transformation between hidden and output layer space is linear (COVER, 1965). It means that the classification of populations nonlinearly separable seems to be linear when the classification is made in a nonlinear multidimensional space (HAYKIN, 1999).

Nonlinear and linear methods of optimization: the use of RBF as an artificial neural networks (ANN) (HAYKIN, 1999) classifier provides some useful properties and capabilities. The capability to separate samples that are not linearly separable is an important property of classifiers because most of the physical applications are based on nonlinear input information. However, a problem with any type ANN is their capability of generalization from a training set to a test set in a real application. Generalization can be influenced by the following factors:

- The architecture of the RBF (HAYKIN, 1999);
- The complexity of the problem (LAU, 1991; HAYKIN, 1999);
- The size of the training set and how representative the set of the considered general population data is (ESPÍRITO SANTO, 2005);
- The methods used to solve the optimization problem.

In this work, we propose a neural network classifier based upon RBF and two methods of solving optimizations problem: nonlinear optimization (Levenberg-Marquart [LM] [WILLIAM et al., 1992]) and combinatorial optimization (simulated annealing [SA]) (KIRKPATRICK; GELATT; VECCHI, 1983). The proposed classifier (RBF-SA) has only one stage, but the network training is carried out in two phases. In the first phase, pre-optimization, one or the entire parameters of the network is estimated using SA. In the second phase, final-optimization, all of the parameters are estimated by LM method. We explore the performance of RBF-SA with different combination of methods of optimization and training procedures. We have carried out experiments in mammographic in order to measure the capability of generalization of the proposed classifier in real applications.

Experiments in mammography: The aim of the experiments with mammography is to investigate the performance of RBF-SA in classifying masses as benign and malignant with edge-sharpness and shape. We describe experiments and present its results in terms of the area under the receive operating characteristic (ROC) curve (METZ, 1986). The classifier is trained and tested with one database consisted of features extracted from 57 regions of interest (ROI) (HILLARY; RANGAYYAN; DESAUTELS, 2003): shape factors (C, F, F_{CC} and SI) and one edge-sharpness measures (A). In order to investigate the influence of two methods of optimizations upon classifier performance, we have made training with no pre-optimization and with pre-optimization of one and two RBF's training parameters. We also have considered several values of structural RBF's parameters during the training, which are conducted in a leave-one-out fashion (HAYKIN, 1999; KIRKPATRICK; GELATT; VECCHI, 1983).

2 Image databases and features

Images: In our simulations we use a database supplied by Departments of Electrical & Computer Engineering and Radiology, University of Calgary, Calgary (AB), Canada. The database is a result of mass and tumor analysis made in mammograms where 57 ROIs are extracted. In each ROI, masses and tumors where manually identified by drawing contours on mammograms, or by consulting radiologist experienced in screening mammography (ANDRÉ; RANGAYYAN, 2003). From the total of 57 contours drawn (ROIs), 37 benign masses and 20 malignant tumors are identified. For each ROI, shape factors (C, F, F_{CC}, SI) and acutance (A) (ANDRÉ; RANGAYYAN, 1990; ESPÍRITO SANTO, 2005) are computed.

Database: The present work does not directly manipulate mammograms. It uses a database (DB) for training the classifier. DB are 57 ROI (37 benign and 20 malignant) result from analysis of masses and tumors made in mammograms where ROIs are manually identified and delimited by expert radiologists in mammographic application. As mentioned before, the information identified on mammogram is used to compute shape factors (C, F, F_{CC}, and SI), and edge-sharpness measure (Acutance–A). To learn details about the ability of the proposed classifier in recognizing benign masses and malignant tumors when a diversity of features, representing multiple characteristics, is used, a combined analysis of all of the features is made. Training a classifier to characterize masses and tumors with a combination of features (shape factor and edge-sharpness, for instance) is more advantageous than training them with shape or edge-sharpness alone (ANDRÉ; RANGAYYAN, 2003). This approach, conducted in (STATISTICAL PACKAGE FOR THE SOCIAL SCIENCES INCORPORATION, 1990), reveals that shape factor of fractional concavity (F_{CC}) and edge-sharpness feature of acutance (A) are meaningful features recommended to classify benign masses and malignant tumor. In addition to feature combination, the training with DB is also performed with leave-one-out procedure (KIRKPATRICK; GELATT; VECCHI, 1983; HAYKIN, 1999).

Feature sets: In the pattern classification experiments, DB is arbitrary discriminated into the following features sets (ESPÍRITO SANTO, 2005):

- S1: Three shape factors (F_{CC}, C and SI) computed from 57 ROI (37 benign and 20 malignant).
- S2: Four shape factors (F_{CC}, C, SI and FF) computed from 57 ROI (37 benign and 20 malignant).
- S3: Four shape factors and edge-sharpness feature (F_{CC}, C, SI, FF and A) computed from 57 ROI (37 benign and 20 malignant).

3 The proposed RBF-SA method

The proposed RBF-SA method is a classifier that discriminates samples (in the present study, mammographic masses and tumors) as malignant or benign. The structural model of proposed classifier is an RBF's network described as:

$$F(x) = \sum_{i=1}^{N} wi \left[\exp\left(-\frac{1}{2\sigma_i^2}\|x - x_i\|^2\right)\right] \quad (1.0)$$

This network consists of a linear combination of multi-variant Gaussian functions with center x_i and standard deviation σ_i. F (x) is the network output for a set of input x_i (input layer); the exponential functions (RBF) are the hidden layer activation functions. Each w_i is the synaptic weight connection between x_i and the output layer via the RBF (HAYKIN, 1999), The training of this network consists of estimating the parameters x_i, σ_i, and w_i regarding the network input vector x.

There are different strategies of training. The choice of a particular type depends on how the centers of RBF are specified. Essentially there are three possibilities (HAYKIN, 1999):

- The centers are fixed and are selected in a random way;
- The centers are selves selected during the supervised training;
- Supervised selection of the centers.

We introduce a new type of RBF's network training where two systems of optimization approaches are used: nonlinear optimization and combinational optimization (HAYKIN, 1999). In this approach some network parameters are estimated in two phases. In the first phase, up to two parameters of the network (x_i and σ_I) are pre-optimized using a combinational optimization method, that is, SA algorithmic (WU, 1993; HAYKIN, 1999). To implement the pre-optimization (pre-estimation) of x_i and σ_I the following parameters are specified:

- T, temperature cooling schedule of the system (KIRKPATRICK; GELATT; VECCHI, 1983; HAYKIN, 1999);
- Fatordec, temperature decay factor (KIRKPATRICK; GELATT; VECCHI, 1983; HAYKIN, 1999);
- Itry, number of Metropolis-Monte Carlo attempt (KIRKPATRICK; GELATT; VECCHI, 1983; HAYKIN, 1999);
- Boltzman criterion (KIRKPATRICK; GELATT; VECCHI, 1983; HAYKIN, 1999).

Typical values used during the pre-optimization are:

- Fatordec = 0.01, 0.1 and 1 for, slow, moderated and fast decay, respectively;
- Itry = 100, 150, 200 etc.

In the second phase, all of the network parameters (x_i, σ_i, and w_i), including those pre-optimized in the first phase, are completely estimated employing a nonlinear optimization approach, such as the LM method (WILLIAM et al., 1992).

4 Experiments and results

Several classification experiments were conducted with up to three sets of features (S1, S2 and S3) and different training strategy with the proposed classifier. In the first strategy, a RBF's network is trained only with LM method (WILLIAM et al., 1992) as parameters optimizer. In the second strategy, a RBF network was trained considering the algorithmic SA as a pre-estimation parameters stage and LM method as a complementary estimation stage of RBF network parameters. From these strategy, where the proposed classifier is now named RBF-SA (KIRKPATRICK; GELATT; VECCHI, 1983), we investigate its performance in classifying benign and malignant features with only pre-estimation of the center of RBF and with pre-estimation of the center and spread matrix.

In order to analyze the influence of the number of features on the performance of RBF-SA,

Table 1: Three sets of features used for training the RBF-SA network

Set	Features
1	F_{CC}, C and SI
2	F_{CC}, C, SI and FF
3	F_{CC}, C, SI, FF and A

Font: Authors.

different combinations of F_{CC}, C, SI, FF, and A, producing the S1, S2 and S3, were used during the training phase. The combinations studied are listed in Table 1.

The experiments conducted can be summarized as follows:

- Train the classifier with the 57 available sets of values of F_{CC}, C, SI, FF, and A, regarding different approach of optimization, using the classifier parameters as shown in Table 2. Repeat until the best performance is obtained (by trial and error), using the leave-one-out procedure (KIRKPATRICK; GELATT; VECCHI, 1983; HAYKIN, 1999). Test the classifier with the same set of features used in the training phase.
- Evaluate the performance of the classifier by using the ROC curve. An ROC curve represents the variation of the true-positive fraction (TPF) versus the false-positive fraction

Table 2: Results (area under the ROC curve A_z in classifying masses as benign or malignant) of RBF-SA using: Factordec = 0.1, total Metropolis-Monte Carlo attempt for, respectively, centers and spread matrix pre-estimating (itry): 100 and 50, k = 0.1, total cooling interaction = 50, T_i = 4, 000 and 3, 12-36, and 1 neurons in the input, hidden, and output layers, respectively.

Feature set	ROC: $A_z \rightarrow$ Optimization method: Only LM	ROC: $A_z \rightarrow$ Optimization method: SA (per-estimation of the centers) + LM (complementary estimation)	ROC: $A_z \rightarrow$ Optimization method: SA (per-estimation of the centers and spread matrix) + LM (complementary estimation)
S1	0.7071	0.9558	0.9776
S2	0.6451	0.9296	0.9459
S3	0.5425	0.8758	0.9045

Font: Authors.

(FPF); the area under the ROC curve (A_Z) may be used as a summarized measure of accuracy (WOODS; BOWYER, 1997; KUPINSKI; ANASTASIO, 1999).

Figure 1 shows the ROC curves obtained after training and testing the RBF-SA network with features set S2. The ROC curves, generated by using Rockit package (KURT ROSSMANN LABORATORIES FOR RADIOLOGIC IMAGE RESEARCH, 2006), shows the classifier performance when only centers and both centers and spread matrix are pre-estimated. During the training phase, the value of the temperature T_i and parameter k are set to 4000 and 0.1, respectively.

The dimension of the space in the hidden layer (M) is 12, 16 and 20 for training with S1, S2 and S3, respectively. The parameter itry (Metropolis-Monte Carlo attempt) has a maximum value of 100 (center pre-estimation) and 50 (spread matrix pre-estimation). The highest classifications accuracies (A_Z = 0.9776, 0.9459 and 0.9045) were provided when the classifier was trained with a combined pre-estimation of the centers and spread matrix.

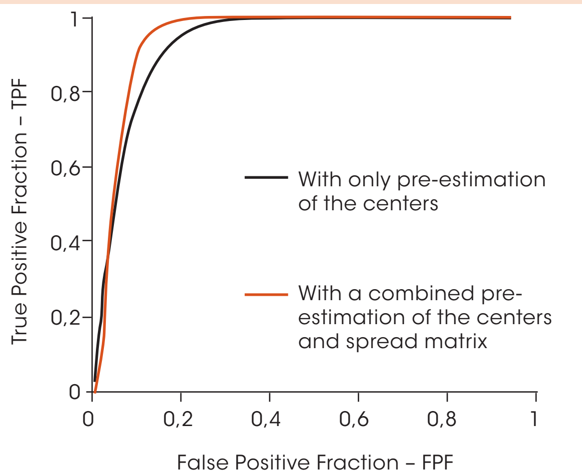

Figure 1: ROC curves for RBF-SA with the features and parameters listed in Table 2

Font: Authors.

5 Discussion

Considering the curves plotted in Figure 1 and the values of A_Z shown in Table 2, the performance of RBF-SA in classifying masses as benign or malignant using features like F_{CC}, C, SI, FF, and A is improved when, at least, one RBF parameter is pre-estimated. As can be seen on Table 2, an increasing in classifier performance is observed from column one to column three. It means that good performance and pre-estimating goes on the same direction. Alto and others (apud HILLARY; RANGAYYAN; DESAUTELS, 2003) have observed that classifying masses and tumors with a multiple features set is better than classifying them with a few ones. However, our experiments apparently show the opposite. We have had a drop in classifier performance as the trainings are conducted from S1 to S3. Yet, such a poor classifier performance is a result of the fall of accuracy of SA algorithmic when the input space dimension is increased. Feature S1, S2 and S3 has dimensions three, four and five, respectively. In order to maintain the same accuracy as it is of training with S1, a boost in the initial cooling temperature (T_i) or an elevation in the number of Metropolis-Monte Carlo attempt should be done for each additional feature. In our experiments T_i and other parameters from SA algorithmic are maintained constantly just because we are interest only in the effect of how sequential pre-estimations can affect the RBF-SA performance. However, good performance in high input dimension is a matter of training parameter starts up. We made additional training with S3 (the worst performance printed out on Table 2) changing only the number of Metropolis-Monte Carlo attempt to 500. In this case, results of A_Z = 0.942 and 0.963 are respectively obtained for column two and three from Table 2.

Yet training under conditions as above tend to be very time consuming. For instance, to produce a RBF-SA performance, in terms of the area under the ROC, printed out in Table 2 (column two and three) a pre-estimating time of 8 hours is needed with S1 using T_i = 4000, 200 Metropolis-Monte Carlo attempt, 50 cooling temperature and M = 12. However, to yield performances such as A_Z = 0.942 and 0.963, as mentioned before, 300 additional iterations are necessary. In this circumstance, the pre-estimating phase takes about 21 hours.

The training parameters (k, T, M, factordec and itry) of the RBF-SA classifier are difficult to be determined in a systematic way. In our simulations, these parameters were experimentally selected. The performance of the classifier depends upon the combination of the values of these parameters and the pre-estimated phases.

As mentioned earlier, the RBF-SA method is not only sensitive to the initial values of the training parameters but also to the number of pre-estimated RFB parameters. Its performance when trained with set features S1 with pre-estimation of center and spread matrix (A_Z = 0.9776) is better than its performance without pre-estimation phases (A_Z = 0.9556) or even without any SA optimization (column one from Table 2).

6 Conclusion

We have presented preliminary results of an investigation on the use of RBF-SA as a classifier mammographic masses into benign or malignant using a combination of shape, edge-sharpness, and pre-estimation phases. High accuracies of 0.9558 and 0.9776 in terms of the area under the ROC curve have been obtained and are comparable to each other to show the influence of pre-estimation on RBF-SA.

References

ANDRÉ, T. C. S. S.; RANGAYYAN, R. M. Classification of tumors and masses in mammograms using neural networks with shape and texture features. In: ANNUAL INTERNATIONAL CONFERENCE OF IEEE ENGINEERING IN MEDICINE AND BIOLOGY SOCIETY, 25., 2003, Cancún. *Paper*... Cancún: IEEE, 2003. n. 6.7.2.4.

BROOMHEAD, D. S.; LOWE, D. Multi-variable functional interpolation and adaptive networks. *Complex Systems*, Champaign, v. 2, n. 3, p. 269-303, 1988.

COVER, T. M. Geometrical and statistical properties of systems of linear inequalities with applications in pattern recognition. *IEEE Transaction on Electronic Computers*, Piscataway, v. EC-14, p. 326-334, 1965.

ESPÍRITO SANTO, R. *Discriminação de classes fundamentada em técnicas de otimização não linear e combinatória associadas a funções de base radiais: um estudo de caso em mamografia*. 2005. Tese (Doutorado em Sistemas Integráveis)-Escola Politécnica, Universidade de São Paulo, São Paulo, 2005.

HAYKIN, S. *Neural network: a comprehensive foundation*. 1. ed. Upper Saddle River: Prentice Hall, 1999.

HILLARY, A.; RANGAYYAN, R. M.; DESAUTELS, J. E. L. An indexed atlas of digital mammograms. In: International Workshop on Digital Mammography, 6., 2002, Bremen. *Proceedings*... Bremen: IWDM, 2003. p. 309-311.

KIRKPATRICK, S.; GELATT, C. D.; VECCHI, M. P. Optimization by simulated annealing. *Science*, Nova York, v. 220, n. 4.598, p. 671-680, 1983.

KUPINSKI, M. A.; ANASTASIO, M. A. Multiobjective genetic optimization of diagnostic classifiers with implication for generating receiver operating characteristic curves. *IEEE Transactions on Medical Imaging*, Piscataway, v. 18, n. 8, p. 675-685, ago. 1999.

KURT ROSSMANN LABORATORIES FOR RADIOLOGIC IMAGE RESEARCH. *Rockit 0.9b beta version*. Chicago: KRL-UofC, 2006. Disponível em: <http://www-radiology.uchicago.edu/krl/KRL_ROC/software_index.htm>. Acesso em: 14 fev. 2006.

LAU, C. G. Y. *Neural networks: theorethical foundation and analyses*. 1. ed. Piscataway: IEEE Press, 1991.

METZ, C. ROC methodology in radiologic imaging. *Invest Radiology*, Temple, v. 21, p. 720-733, set. 1986.

MOODY, J.; DARKEN, C. J. Fast learning in networks of locally-turned processing units. *Neural Computation*, San Diego, v. 1, p. 281-294, 1989.

POGGIO, T.; GIROSI F. Regularization algorithms for learning the are equivalent to multi-layer networks. *Science*, Nova York, v. 247, p. 978-982, 1990.

POWELL, M. J. D. Radial basis functions for multivariable interpolations: a review. In: IMA CONFERENCE ON ALGORITHMS FOR THE APPROXIMATIONS OF FUNCTIONS AND DATA, 1985, Shrivenham. *Annals*... Shrivenham: RMCS, 1985. p. 143-167.

RENALS, S. Radial basis functions network for speech pattern classification. *Electronics Letters*, Piscataway, v. 25, p. 437-439, 1989.

STATISTICAL PACKAGE FOR THE SOCIAL SCIENCES INCORPORATION. *Advanced statistics user's guide*. Version 10. Chicago: SPSS, 1990.

WILLIAM, H. P. et al. *Numerical recipes in c: the art of scientific computing*. 1. ed. Cambridge: Cambridge University Press, 1992.

WOODS, K.; BOWYER, K. W. Generating ROC curves for artificial neural networks. . *IEEE Transactions on Medical Imaging*, Piscataway, v. 16, n. 3, p. 329-337, 1997.

WU, Y. Artificial neural networks in mammography: application to decision making in the diagnosis of breast cancer. *Radiology*, Richmond, v. 187, p. 81-87, abr. 1993.

Recebido em 14 fev. 2006 / aprovado em 3 abr. 2006

Para referenciar este texto

ESPÍRITO SANTO, R. do; LOPES, R. de. D.; RANGAYYAN, R. M. Classification of mammographic features using RBF-SA. *Exacta*, São Paulo, v. 4, n. 2, p. 317-324, jul./dez. 2006.