A genetic algorithm based nearest neighbor classification to breast cancer diagnosis

R. Jain[1] and J. Mazumdar[2]

[1]*School of Information Technology, James Cook University*
[2]*Dept of Applied Mathematics, The University of Adelaide, South Australia*

Abstract

This paper presents an application of a hybrid approach (the genetic algorithms and the k-nearest neighbour) proposed by Ishbuchi [10] to *Wisconsin breast cancer data*. For the diagnosis of breast cancer, the determination of the presence of *benign/malignant* breast tumors represents a very complex problem (even for an experienced cytologist) [4]. Therefore the automatic classification of *benign* and *malignant* symptoms is highly desirable as a valuable aid to assist oncologists in the decision making of the diagnosis of breast cancer. In this paper, the genetic algorithm based k-nearest neighbour method for classification of benign and malignant breast tumors is presented. The genetic-algorithm (GA) is used for finding a compact reference set by selecting a small number of reference patterns from a large number of training patterns in nearest neighbor classification. The GA simultaneously performs feature selection and pattern selection and prunes unnecessary features. The goal is to maximize the classification performance of the reference set and minimize the number of selected patterns and features. Results are also compared with a fuzzy-genetic approach where each reference pattern represents a fuzzy if-then rule with a circular-cone-type membership function.

Introduction

Breast cancer is the most common cancer in women in many countries. Most breast cancers are detected as a lump/mass on the breast, by self-examination/mammography, or by both [7]. Screening mammography is the best tool available for detecting cancerous lesions before clinical symptoms appear. Surgery - either a biopsy or a lumpectomy have been the most common method to remove them. Fine needle aspiration (FNA) of breast masses is a cost-effective, non-traumatic, and mostly non-invasive diagnostic test that obtains information needed to evaluate malignancy. Recently, a new minimally invasive technique, which uses super-cooled nitrogen to freeze and shrink a non-cancerous tumor and destroy the blood vessels feeding the growth of the tumor has been developed [7] in USA. With increased interpretability in mind as a prior objective, a number of researchers have applied the method of extracting Boolean rules from neural networks. Their results are encouraging, exhibiting both good performance and a reduced number of rules and relevant input variables.

Corresponding author: R. Jain, School of Information Technology, James Cook University, Tel: (07) 4042 1211, Fax: (07) 4042 1284, Email: ravi.jain@jcu.edu.au

Artificial Intelligence techniques (such as neural networks, fuzzy logic, genetic programming and combination of these methods) have attracted a lot of attention in the area of medical diagnosis [5], [9], [20]. These techniques are successfully applied to a wide variety of decision-making problems.

Nearest neighbor classification [5] is one of the most well-known non-parametric classification techniques. Their roots lie in the fields of statistics and pattern recognition. *Non-parametric decision rules*, such as the *nearest neighbour rule*, are attractive because no prior knowledge of the distributions is required. These rules rely on the membership of unknown objects instead on the training set of objects with known class membership to make decisions. The nearest neighbour rule, classifies an unknown object to the class of its nearest neighbour in the measurement space using, most commonly, Euclidean metrics. Despite being extremely simple, the method is not widely used to solve pattern recognition problems because of the fact that they consume a large memory space and long computing time. This issue may not be applicable at present because of availability of fast computers and cheap hardware.

In this paper, nearest neighbour optimisation is achieved by *genetic* algorithms (GA). A genetic algorithm simultaneously performs feature selection and pattern selection. The Nearest Neighbors Classifier

In the basic nearest neighbour rule classifier, each training sample described by their features is used as a prototype and a test sample is assigned to the class of the closest prototype[1]. The k-nearest neighbors performs vote

on the class of a new sample based on the classes of the k nearest training samples. Closeness is usually defined in terms of a metric distance on the Euclidean space with the input measurement variables as axes. It attempts to improve upon the previous rule. Its asymptotic classification error is bounded above by twice the Bayes error.

Let **x** be the feature vector for the unknown input, and let $(n_1, n_2, ..., n_m)$ be the training patterns for the c classes. Then the error in matching x against n_i, is given by:

$$\|\mathbf{x} - n_i\| \tag{1}$$

The Euclidean distance metric is commonly used to calculate the k-nearest neighbors. $D(x, x^i)$ denote the Euclidean distance between two patterns vectors, **x** and n, then

$$D(\mathbf{x}, n_i) = \left[\sum_{j=1}^{d} (x - n_i)^2 \right]^{\frac{1}{2}} \tag{2}$$

and d is the number of features. $D(\mathbf{x}, n_i)$ as the matching score between the test patterns x and the training pattern n_i. So finding the minimum Euclidean distance is equivalent to finding the maximum matching score. If the class of the nearest neighbor x_p is the same as that of the new pattern x, is correctly classified. Otherwise the new pattern x is misclassified.

In this paper, the goal is to maximize the number of correctly classified training patterns by the reference set and minimize the number of selected patterns and features. In the genetic algorithm, a reference set is coded by a binary string of the length (x + y). Each bit value of the first x bits represents the inclusion/exclusion of the corresponding feature. The other y bit values represent the inclusion/ the exclusion of the given x training patterns. A fitness value of each reference set is defined by the weighted sum of the correctly classified training patterns by the reference set plus the number of selected patterns and features. Simulation results show that a small number of training patterns are selected by this approach together with a few important features. Finally an extension of this approach to the case where each reference pattern has its own localized receptive field and strength. The radius of the localized receptive field of each reference pattern is automatically specified in a heuristic manner from the relation with the other reference patterns. The strength is determined by training patterns in the localized receptive field. In this extended approach, each reference pattern can be viewed as a fuzzy if-then rule with a circular-cone-type membership function.

$$\mu_p(x) = \max \left\{ 0, 1 - \frac{d(\mathbf{x}_p, \mathbf{x})}{r_p} \right\} \tag{3}$$

where $\mathbf{x}_p$ and r_p are the center and the radius respectively of the membership function of the fuzzy set $\mathbf{A}_p$.
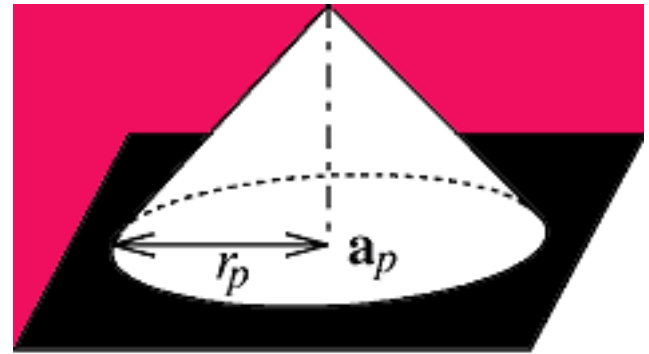

Figure 1. *A circular-cone-type membership function.*

Genetic algorithms

Genetic Algorithms (GA) are search algorithms that use operations found in natural genetics to guide the trek through a search space. GA use a direct analogy of natural behavior. They work with a population of individuals. Each representing a possible solution to a given problem. Each individual is assigned a fitness score according to how good a solution to the problem it is. GA is theoretically and empirically proven to provide robust search in complex spaces giving a valid approach to problems requiring efficient and effective search.

GA starts with a population of randomly generated solutions chromosomes and advances toward better solutions by applying genetic operators modeled on the genetic processes occurring in nature. In these algorithms we maintain a population of solutions for a given problem this population undergoes evolution in a form of natural selection. In each generation relatively good solutions reproduce to give offspring that replace the relatively bad solution which die. An evaluation or fitness function plays the role of the environment to distinguish between good and bad solutions. Although there are many possible variants of the basic GA the fundamental underlying mechanism operates on a population of chromosomes or individuals representing possible solutions to the problem and consists of three operations

- evaluation of individual fitness
- formation of a gene pool and
- recombination using two basic genetic operators crossover and mutation.

The figure shows the structure of a simple GA.

To run genetic algorithms for pattern and feature selection problems, a reference set (RS) = R*S(F, P)* of the length (x + y) is coded as a binary string (BS)

$$BS = x_1 x_2 \dots x_x \, p_1 p_2 \dots p_y, \tag{4}$$

where the first x bits represents the inclusion/the exclusion of the *n* features, and the other y bits represents the inclusion or the exclusion of the m patterns.

The feature set *F* and the pattern set P are obtained by decoding the string S as follows:

=0	set generation number to zero
initpopulation P(0)	initialise a usually random population of individuals
evaluate P(0)	evaluate fitness of all initial individuals of population
while (not done) do	test for termination criterion (time, fitness, etc.)
Begin	
i = i + 1	increase the generation number
select P(i) from P(i-1)	select a sub-population for offspring reproduction
recombine P(i)	recombine the genes of selected parents
Mutate P(i)	perturb the mated population stochastically
evaluate P(i)	evaluate its new fitness
End	

Figure 2. *Pseudo code of a genetic algorithm.*

$$F = \left\{ f_i \middle| a_{i=1,} i = 1, 2, ..., n \right\} \tag{5}$$

$$F = \left\{ \mathbf{x}_p \middle| s_p = 1, i = 1, 2, ..., m \right\} \tag{6}$$

In order to maximise the number of correctly classified patterns (NCP) and minimize number of features and patterns the following fitness function for the binary string (BS) defined in (4) was devised:

$$fitness(RS) = WCP.[\#CP(RS)] - WF.[\#F] - P.[\#P] \tag{7}$$

where, #CP(RS) represents the number of correctly classified patterns by reference set (RS), #F represents the number of features in F.
#P represents the number of patterns in P.
WNCP, WF, and WP represent positive weights.
where WNCp, WF and Wp are non-negative weights.
    In genetic algorithm,
    A.   a number of binary strings of the length (x + y) are randomly generated to form an initial population. Let us denote the population size by *Ps*.
    B.   a pair of strings are randomly selected from the current population. Two strings are generated by crossover and mutation. The selection, crossover and mutation are iterated

to generate *Ps* strings. The newly generated Ps strings are added to the current population to form the enlarged population of the size $2.Ps$. The next population is constructed by selecting the best *Ps* strings from the enlarged population. The population update is iterated until a pre-specified stopping condition is satisfied.
    In genetic algorithm, the uniform crossover was applied to avoid the dependency of the performance on the order of *n* features and *m* patterns in the string. For efficiently decreasing the number of reference patterns, the biased mutation was used where a larger probability is assigned to the mutation from " $s_p = 1$ " to " $s_p = 0$ " than the mutation from " $S_P = 0$ " to " $S_P = 1$ ". That is, two different mutation probabilities $p_x(1 \to 0)$ and $p_x(0 \to 1)$ for the last x bits of the string, each of which represents the inclusion or the exclusion of the corresponding pattern in the reference set. Since $p_x(0 \to 1) < p_x(1 \to 0)$, the number of reference patterns is efficiently decreased by the biased mutation during the execution of genetic algorithm. The biased mutation is the main characteristic feature of genetic algorithm. The biased mutation used because the number of selected reference patterns is to be much smaller than that of the given patterns. It has been shown in [10] that the number of reference patterns could not be efficiently decreased without the biased mutation. Note that use of the standard unbiased mutation for the first y bits of the string, each of which represents the inclusion or the exclusion of the corresponding feature. This is because usually the number of given features is not as large as that of given patterns (i.e., usually the biased mutation is not necessary for the feature selection).

Fuzzy nearest neighbour classification
    In this approach, each reference pattern has a different strength and a different receptive field. Each reference pattern is handled as the following fuzzy if-then rule:

If x is AP then Class with CFp, (8)

where A_P is a fuzzy set with a circular-cone-type membership function, C_p is the consequent class, and CF_p is the strength of the fuzzy if-then rule. For the case of the two-dimensional pattern space the center of the fuzzy set A_P is the same as the reference pattern x_p in the circular-cone-type membership function of the fuzzy set A_P in Fig. The radius r_p of the localized receptive field of the reference pattern x_p is determined by the relation with the other reference patterns as follows:

$$r_p = \alpha . \min \left\{ d\left(x_p x_q \right) \middle| x_q \in P, q \neq p \right\} \tag{9}$$

where α is a positive constant, $d(\mathbf{x_p}, \mathbf{x_q})$ is the Euclidean distance between x_P and x_q and P is the set of reference patterns. When $r_p = 0$, a very small value to r_p is assigned. In computer simulations, r_p set as $rp = 0.00001$ when r_p was calculated as $r_p = 0$ in (9). As shown in (9), the radius

r_p is determined by the distance to the nearest reference pattern. When reference patterns are sparsely distributed in the pattern space, each reference pattern has a large localized receptive field. On the contrary, when there are many reference patterns in a small area of the pattern space, each reference pattern has a small localized receptive field. In computer simulation, we specified the constant α as $\alpha = 2$.

The consequent class C_p of the fuzzy if-then rule is the same as the class label of the reference pattern x_p. The strength CF_p is specified in a heuristic manner from the training pattern in the localized receptive field of x_p (i.e., the subset of the pattern space where the membership function of A_P assumes a positive value). For specifying the strength CF_p, we first calculate the sum of the compatibility grades of the training patterns from each class as

$$\beta_{class\ h} = \sum_{x_p \in class\ h} \mu_{j1}(x_{p1}) \dots \mu_{jn}(x_{pn}), \quad h = 1,2,\dots,c \qquad (10)$$

where c is the number of given classes. Then we specify CF_p as

$$CF_j = \frac{(\beta_{class\ \hat{h}} - \bar{\beta})}{\sum_{h=1}^{c} \beta_{class\ h}} \qquad (11)$$

When CF_p is negative, the reference pattern X_P is

where

$$\bar{\beta} = \sum_{h \neq \hat{h}} \frac{\beta_{class\ h}}{(c-1)}$$

removed. A new pattern x is classified by a single winner rule, which is specified as

$$\mu_j(x).CF_j = \max\left\{ \mu_p(x).CF_p \middle| x_p \in P \right\} \qquad (12)$$

Wisconsin diagnostic breast cancer data

The Wisconsin *breast cancer* dataset was obtained from repository of machine learning database University of California, Irvine. This data set has 32 attributes (30 real-valued input features) and 569 instances of which 357 benign and 212 malignant class. There are several studies based on this database. Bennet and Mangasarian [2] used linear programming techniques, obtaining a 99.6% classification rate on 487 cases (the reduced database available at the time). However, diagnostic decisions are essentially black boxes, with no explanation as to how they were attained.

Ten real-valued features are computed for each cell nucleus are based on fine needle aspirate analysis:

a) radius (mean of distances from center to points on the perimeter)
b) texture (standard deviation of gray-scale values)
c) perimeter
d) area
e) smoothness (local variation in radius lengths)
f) compactness (perimeter^2 / area - 1.0)

Random seed (pa.seed)	10
number of population (pa.pop)	100
Probability of 1 when initializing (pa.ini_1)	0.5
weight for classification (pa.w_ncp)	10.0
weight for the number of rules	5.0
weight for the number of attributes	1.0
mutation rate from 0 to 1 (pa.m_0 to1)	0.01
mutation rate from 1 to 0 (pa.m_1to 0)	0.1
mutation rate for attributes	0.01
number of generations (pa.gen)	1000
k-nearest neighbor (pa.k)	1
crossover probability (pa.c_o)	1.0

Table 1. *GA parameters for nearest neighbour rule.*

Random seed (pa.seed)	10
Number of population (pa.pop)	100
Rate of radius of the rule (pattern)pa.r_rate	2.0
probability of 1 when initializing (pa.ini_1)	0.5
weight for NCP (pa.w_ncp)	4.0
weight for the number of rules	1.0
weight for the number of attributes	1.0
mutation rate from 0 to 1 (pa.m_0to1)	0.01
mutation rate from 1 to 0 (pa.m_1to0)	0.1
mutation rate for attributes	0.01
Number of generations (pa.gen)	100
Number of learning of cf (pa.learn)	0
Learning rate eta_1 (pa.eta_1)	0.001
Learning rate eta_2 (pa.eta_2)	0.1
crossover probability (pa.c_o)	1.0

Table 2. *GA parameters for fuzzy genetic approach.*

The genetic-algorithm-based feature and pattern selection method with the above modification to the data were applied. After the 1000 generations 9 patterns were selected by the genetic algorithms. The main advantage of the above fuzzification of nearest neighbor classification is that the radius of the localized receptive field r_p and the strength CF_p are assigned to each reference pattern x_p. From these two values, we can understand the typicality and the importance of each reference pattern. The discussion on the effect of the above fuzzification on the performance of nearest neighbor classification needs more computer simulations.

Using non-fuzzy approach 558 of 569 patterns with 9 of 32 features were correctly classified and 11 patterns were

Australas. Phys. Eng. Sci. Med. Vol. 26, No 1, 2003

Jain *et al* ● A genetic algorithm

misclassified using 33 rules as shown in Table 3. Using fuzzy-genetic approaches 523 of 569 patterns were correctly classified and 46 patterns were misclassified using 9 rules as shown in Table 4.

# Correctly classified patterns	# Misclassified patterns	Rules
558	11	33

Table 3. *Results of Non-fuzzy approach.*

# Correctly classified patterns	# Misclassified patterns	Rules
523	46	9

Table 4. *Results of fuzzy-genetic approach.*

Conclusion

In this paper we combined the genetic algorithms and the k-nearest neighbour approach to the design of compact reference sets for the classification of Wisconsin breast cancer data. They attain high classification performance and the resulting systems involve a few simple rules, and are therefore interpretable. Feature selection and pattern selection are simultaneously performed by a genetic algorithm. That is, a small number of reference patterns with only important features are selected. An extension of this approach to the case where each reference pattern was handled as a fuzzy if-then rule with a circular-cone-type membership function was also compared. This extension could reduce the number of reference patterns with no deterioration of the classification performance.

Acknowledgements

References

1. Cover, T. M., Hart, P. E. *Nearest Neighbor Pattern Classification",* IEEE Transactions on Information Theory, vol. IT-13, No.1, 1967, pp.21-27

2. Duda, R O., Hart, P. E. *Pattern Classification and Scene Analysis,* John Wiley & Sons, New York, 1973.

3. Dasarathy, B. V. (ed.), *Nearest Neighbour (NN) Norms: NN Pattern Classification Techniques*, IEEE Computer Society Press, 1991.

4. Salfity, M. F., Kaufmann, G. H., Granitto, P. and Ceccatto, H. A. *A Computer-Aided Diagnosis Method for Automated Detection and Classification of Clustered Microcalcifications in Mammograms* 3er. Simposio Argentino de Informatica y Salud – SADIO.

5. DeSilva, C. J. S. et al. *Artificial Neural networks and Breast Cancer Prognosis.* The Australian Computer Journal ,26, 78-81,1994.

6. Uebele, V. et al. "*A Neural Network-Based Fuzzy Classifier*" IEEE trans. on Systems, Man, and Cybernetics Vol. 25, No. 2, pp. 353-361, (1995).

7. "*The Weekend Australian*" July13-14, 2002, Health Section, Page 7.

8. Cios, K. J. et. al. "*Using Fuzzy Sets to Diagnose Coronary Artery Stenois*", IEEE Computer, 57-63, (March 1991).

9. Watanabe, H. et. al. "*The Application of a Fuzzy Discrimination Analysis for diagnosis of Valvular Heart Disease*" IEEE trans. on Fuzzy Systems, (1994).

10. Ishibuchi, H. et.al "*Pattern and Feature Selection By Genetic Algorithms in Nearest Neighbour Classification*" unpublished report.

11. Hayashi, Y. "*Neural Expert System Using Fuzzy Teaching Input and Its Application to Medical Diagnosis*", Proceedings of the Second International Conference on Fuzzy Logic and Neural Networks, Iizuka, Japan, 989-993, (July 1992).

12. Arita, S. et. al. "*Supporting System for Medical Diagnosis Using Natural Language by a Fuzzy Inference*", Proceedings of the Second International Conference on Fuzzy Logic and Neural Networks, Iizuka, Japan, 1017-1019, (July 1992).

13. Arita, S. et. al. "*Multistage Process of Fuzzy Inference and its Application to the Medical Diagnosis*", Proceedings of the Third International Conference on Fuzzy Logic, Neural Nets and Soft Computing, Iizuka, Japan, 323-324, (August 1994).

14. Esogbue, A. O. "*Measurements and Valuation of a Fuzzy Mathematical model for Medical diagnosis*" Fuzzy Sets and Systems, pp 223-242, (1983).

15. Giove, S. et. al. "*An Adaptive Fuzzy Control Module for Automatic Dialysis", Proceedings of the 8th Austrian AI Conference, FLAI''93,* Linz, Austria, Edited by E.P. Klement and W. Slany, Springer-Verlag, (June 1993).

16. Giove, S. et. al. "*An Adaptive Fuzzy Control Module for Automatic Dialysis*", Proceedings of the 8th Austrian AI Conference, FLAI'93, Linz, Austria, Edited by E.P. Klement and W. Slany, Springer-Verlag, (June 1993).

17. Hall, L. O. et. al. "*A Comparison of Neural Network and Fuzzy Clustering Techniques in Segmenting Magnetic Resonance Images of the Brain*", IEEE Transactions on Neural Networks, 3(5), 672-682, (September 1992).

18. Hilera, J. R. et. al. "*Neural Networks for ECG Compression and Classification*", Proceedings of the Third International Conference on Fuzzy Logic, Neural Nets and Soft Computing, Iizuka, Japan, 121-124, (August 1994).

19. Ichimura, T. et. al. "*Medical Diagnosis Using a Parallel Fuzzy Inference Performed by Neural Networks*", Proceedings of the Second International Conference on Fuzzy Logic and Neural Networks, Iizuka, Japan, 1003-1006, (July 1992).

20. Ishibuchi, H. et. al, "*A fuzzy classifier system that generates fuzzy if-then rules for pattern classification problems,*" Proc. Int. Conf. Evolutionary Computat. Perth, Australia, Nov. 1995, vol. 2, pp 759-764.

21. Karr, C. L. "*Fuzzy control of an exothermic reaction using genetic algorithms*", Engineering applications of artificial intelligence, Vol. 6, No. 6, pp. 575-582, (1993).

22. Lee, C.C., "*Fuzzy logic in control systems: fuzzy logic controller -Part 1*, IEEE Transactions on Systems, Man, and Cybernetics, Vol. 20, No. 2, pp. 404-418, (March/April 1990).

23. Makino, K., et.al. "*An Evaluation of hoarseness by the Fuzzy Logic*", Proceedings of the Third International Conference on

Australas. Phys. Eng. Sci. Med. Vol. 26, No 1, 2003

Jain *et al* ● A genetic algorithm

Fuzzy Logic, Neural Nets and Soft Computing, Iizuka, Japan, 317-318, (August 1994).

24. Makino, M., et.al. "*Development of Fuzzy Logic Program ProFINa and Application to Adverse Drug Effect Monitoring*", Proceedings of the Third International Conference on Fuzzy Logic, Neural Nets and Soft Computing, Iizuka, Japan, 311-312, (August 1994).

25. Meier, R. et.al. "*Fuzzy Logic Control of Blood Pressure During Anesthesia*", IEEE Control Systems Magazine, 12-17, (December 1992).

26. Norris, D., et.al. "*Medical Diagnosis from patient Records - A method using fuzzy Discrimination and Connectivity Analysis*" in Fuzzy Sets and Systems, pp. 73 - 87, (1987).

27. Shigeo, A., et al.,"*Tuning of a Fuzzy Classifier derived from data*" International trans. on Fuzzy Systems, Vol. 5, No. 3, pp. 358-368, (1997).

28. Yoshizawa, et.al. "*An Automatic Monitoring and Estimation Tool for the Cardiovascular System Under Ventricular Assistance Using Fuzzy Reasoning*", Proceedings of the Second International Conference on Fuzzy Logic and Neural Networks, Iizuka, Japan, 1013-1016, (July 1992).

29. Zadeh, L.A., *Fuzzy Logic* IEEE Computer, pp. 83-93 (April 1988).

30. Mendel, J. M. "*Fuzzy logic systems for engineering: A tutorial.* Proceedings of the IEEE, 83(3): 345-377, March 1995.

31. Yager R. R. and Zadeh, L. A. *Fuzzy Sets, Neural Networks, and Soft Computing.* Van Nostrand Reinhold, New York, 1994.