DEVELOPMENT OF A BREAST CANCER DIAGNOSTIC SYSTEM VIA A FUZZY-GENETIC SCHEME

V.S. KODOGIANNIS

13-99

Preprint no. 13-99/1999

Department of Computer Science
University of Ioannina
451 10 Ioannina, Greece

Development of a breast cancer diagnostic system via a fuzzy-genetic scheme.

V.S. Kodogiannis

Department of Computer Science, University of Ioannina, Ioannina, GR-45110, Greece,

Tel: +30-651-97308, Fax: +30-651-48131, Email: vassilis@cs.uoi.gr

ABSTRACT: The presentation of an evolutionary approach for discovering fuzzy systems for breast cancer diagnosis is considered in this study. The automatic diagnosis of breast cancer is an important real-world problem. In this study, we focus on the Wisconsin breast cancer diagnosis (WBCD) problem, combining two schemes, fuzzy systems and evolutionary algorithms, so as to automatically produce diagnostic systems. The proposed fuzzy-genetic approach produces systems exhibiting one of the highest performances shown to date as well as the possibility of attributing a confidence measure to the output diagnosis. In addition, the few simple rules involved in the resulting system make it also human-interpretable.

Key words: Fuzzy systems, genetic algorithms, breast cancer diagnosis.

1. INTRODUCTION

A major class of problems in medical science involves the diagnosis of disease, based upon various tests performed upon the patient. When several tests are involved, the ultimate diagnosis may be difficult to obtain, even for a medical expert. This has given rise, over the past few decades, to computerised diagnostic tools, intended to aid the physician in making sense out of the welter of data. A prime target for such computerised tools is in the domain of cancer diagnosis. Specifically, where breast cancer is concerned, the treating physician is interested in ascertaining whether the patient under examination exhibits the symptoms of a benign case, or whether her case is a malignant case.

An efficient computerised diagnostic tool should possess two characteristics, which are often in conflict. First, the tool must attain the highest possible performance, *i.e.,* diagnose the presented cases correctly as being either *benign* or *malignant*. Moreover, it would be highly desirable to be in possession of a so-called *degree of confidence*: the system not only provides a binary diagnosis (*benign* or *malignant*), but also outputs a numeric value which represents the degree to which the system is confident about its response. Second, it would be highly beneficial for such a diagnostic systems to be human-friendly, exhibiting so-called interpretability. This means that the physician is not faced with a black box that simply spouts answers (albeit correct) with no explanation; rather, we would like for the system to provide some insight as to how it derives its outputs.

In this study, two schemes, fuzzy systems and evolutionary algorithms, are combined to automatically produce systems for breast cancer diagnosis. The major advantage of fuzzy systems is that they favour interpretability, however, finding good fuzzy systems can be quite an arduous task. This is where evolutionary algorithms step in, enabling the automatic production of fuzzy systems, based on a database of training cases.

Fuzzy modelling is the task of identifying the parameters of a fuzzy inference system so that a desired behaviour is attained. The parameters of fuzzy inference systems can be classified into four categories: logical, structural, connective and operational. On the other hand, evolutionary algorithms are used to search large, and often complex, search spaces. They have proven worthwhile on numerous diverse problems, able to find near-optimal solutions given an adequate performance (fitness) measure. Fuzzy modelling can be considered as an optimisation process where part or all of the parameters of a fuzzy system constitute the search space. Evolutionary fuzzy modelling although had first appeared about a decade ago, it has been applied in an ever-growing number of domains, like chemistry, medicine, telecommunications and of course control.

Both connective and structural parameters modelling can be viewed as rule base learning processes with different levels of complexity. In the evolutionary algorithm community there are two major

approaches for evolving such rule systems: The Michigan approach and the Pittsburgh approach [1]. A more recent method has been proposed specifically for fuzzy modelling: the iterative rule learning approach [2]. This study is organised as follows: In the next section, the Breast cancer diagnosis problem, based on the Wisconsin Hospital Database, is presented. In section 3, the proposed evolutionary approach to the WBCD problem is described, while the results are discussed in section 4, followed by concluding remarks in section 5.

2. THE BREAST CANCER DIAGNOSIS PROBLEM

Breast cancer is the most common cancer among women and a frequent cause of death in the 35-55 age group. The presence of a breast mass is am alert sign, but it does not always indicate a malignant cancer. 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. The Wisconsin breast cancer diagnosis (WBCD) database [3] is the result of the efforts made at the Univ. of Wisconsin Hospital for accurately diagnosing breast masses based solely on an FNA test [4]. Nine visually assessed characteristics of an FNA sample considered relevant for diagnosis were identified, and assigned an integer value between 1 and 10. The measured variables are illustrated as follows:

- Clump Thickness *(v1)*
- Uniformity of Cell Size *(v2)*
- Uniformity of Cell Shape *(v3)*
- Marginal Adhesion *(v4)*
- Single Epithelial Cell Size *(v5)*
- Bare Nuclei *(v6)*
- Bland Chromatin *(v7)*
- Normal Nucleoli *(v8)*

- Mitosis *(v9)*

Specialists in the field furnished the diagnostics in the WBCD database. The database itself contains 683 cases, with each entry representing the classification for a certain ensemble of measured values. However, the diagnostics do not provide any information about the degree of benignity or malignancy. There are several studies based on this database. Using the C4.5, the well-known method for decision tree induction, a classification rate of 95.7% has been obtained [5]. In another similar technique, the OC1, a method for the induction of oblique decision trees, a rate of 96.2% has been achieved [6]. Kermani *et al.* [7] used a genetic algorithm to extract the most important variables, their attained performance level being lower, 94.7%. Setiono [8] proposed a method based on pruned neural networks for finding a set of rules to explain the diagnostic. His results were encouraging, exhibiting both good performance, and a reduced number of rules and relevant input variables, 95.42% with 1 rule per system and 97.14% with 3 rules per system. However, the extraction of rules was a manual, experience-based process. Finally, the well-known neuro-fuzzy NEFCLASS learning algorithms have been applied to the above-mentioned problem [9]. After training the fuzzy sets, with four rules, for 100 epochs the classification accuracy reached at 96.5%, *i.e.* 24 classification errors. The work in this study is focused on a two-component set-up, in order to evolve fuzzy rules for the WBCD problem: the fuzzy inference system itself and the genetic algorithm.

3. EVOLUTIONARY FUZZY SYSTEMS FOR THE WBCD PROBLEM

The proposed scheme for the WBCD problem is illustrated in Fig. 1. It consists of a fuzzy system and a threshold unit. The fuzzy system computes a continuous appraisal value of the malignancy of a case, based on the input values. The threshold unit then outputs a *benign* or *malignant* diagnostic according to the fuzzy system's output.

3.1 Fuzzy system parameters

Previous knowledge about the WBCD problem and about some of the extant rule-based models represents valuable information to be used for the choice of fuzzy parameters. However, the following issues should be taken into consideration: small number of rules [10], and small number of variables [10][11].

Some fuzzy models forgo interpretability in the interest of improved performance. Where medical diagnosis is concerned, interpretability is the major advantage of fuzzy systems. This motivated us to take into account the following five semantic criteria, defining constraints on the fuzzy parameters [12], Distinguish-ability, justifiable number of elements, coverage, normalisation and orthogonality. Taking in account the above criteria, the fuzzy system setup is constructed as:

Logic parameters

- Reasoning mechanism: zero-order, Takagi-Sugeno-Kang (TSK) fuzzy system
- Fuzzy operators: min and max
- Input membership function type: orthogonal, trapezoidal as shown in Fig. 2.

Structural parameters

- Relevant variables: there is insufficient a priori knowledge to define them, therefore this will be one of the genetic algorithm's goals.
- Number of membership functions: two membership functions denoted, *Low* and *High* are used for the input variables. Three membership functions have been used but the results were less satisfactory, probably due in part to the increased search space size. Two output values are used, corresponding to b*enign* and, m*alignant* diagnostics.
- Number of rules: results from Setiono [8] show that few rules are needed to achieve good performance. In this study the number of rules was limited to be between 1-5. These rules are evolved.

Connection parameters

- Antecedents of rules: to be found by evolution.

- Consequent of rules: the implemented strategy has the algorithm find rules for one of the possible consequent (*malignant* or *benign*), the other being an ***else*** condition.
- Rule weights: active rules have a weight of value 1 and the ***else*** condition has a weight of 0.25.

Operational parameters

- Input membership function values: to be found by evolution.
- Output membership function values: following the WBCD database, we used a value of 2 for *benign* and 4 for *malignant*.

3.2 The Genetic Algorithm

The Pittsburgh-style structure learning is applied [1], using a genetic algorithm to search for three parameters. The genome, encoding relevant variables, input membership function values, and antecedents of rules, is constructed as follows:

- *Membership function parameters:* There are nine variables (*v1-v9*), each with two parameters *P* and *d*, defining the start point and the length of the membership function edges, respectively as shown in Fig. 2.
- *Antecedents:* The i^{th} rule has the form:

  if (*v1* **is** A_1^i) and ...and (*v9* is A_9^i) **then** (*output* **is** *benign*),

  where A_j^i represents the membership function applicable to variable v_j. A_j^i can take on the values:

  1 (*Low*), 2 (*High*), or 0 or 3 (*Other*).
- Relevant variables are searched for implicitly by letting the algorithm choose non-existent membership functions as valid antecedents; in such a case the respective variable is considered irrelevant.

Table 1 delineates the parameters encoding, which together form a single individual's genome. Fig. 3 shows a sample genome.

To evolve the fuzzy inference systems, we have applied a standard genetic algorithm [1] with a fixed population size of 200 individuals, and fitness-proportionate selection. The algorithm terminates when the maximum number of generations, G_{max}, is reached (we set $G_{\max} = 2000 + 500 \times N_r$), *i.e.* dependent on the number of rules N_r used in the run), or when the increase in fitness of the best individual over five successive generations falls below a certain threshold (in the experiments, threshold values between 2×10^{-7} and 4×10^{-6} have been used).

The proposed fitness function combines three criteria:

- F_c: classification performance, computed as the percentage of cases correctly classified;
- F_e: the quadratic difference between the continuous appraisal value (in the range [2,4]) and the correct discrete diagnosis given by the WBCD database (either 2 or 4);
- F_v: the average number of variables per active rule. The fitness function is given by $F = F_c - \alpha F_v - \beta F_e$, where α=0.05 and β=0.01 (these latter values were derived empirically).

F_c, the percentage of correctly diagnosed cases, is the most important measure of performance. F_v measures the linguistic integrity (interpretability), penalising systems with a large number of variables per rule (on average). F_e adds selection pressure towards systems with low quadratic error.

4. RESULTS

This section describes the results obtained when applying the methodology described in the previous section. We first delineate the success statistics relating to the evolutionary algorithm. Then we describe in full three evolved fuzzy systems, a 3–rule system, a 2-rule system and a 1-rule system that exemplify our approach. Finally we discuss the issue of obtaining a confidence measure of the system's output, going beyond a mere binary, *benign-malignant* classification.

4.1 The genetic approach

The evolutionary experiments performed fall into three categories, in accordance with the data repartitioning into two distinct sets: training set and test set. The three categories are:

- training set contains all 683 cases of the WBCD database, while the test set is empty,
- training set contains 75% of the WBCD cases, and the test set contains the remaining 25% of the cases,
- training set contains 50% of the WBCD cases, and the test set contains the remaining 50% of the cases,

A total of 120 evolutionary runs were performed, all of which found systems whose classification performance exceeds 94.5%. In particular, considering the best individual per run (*i.e.* the evolved system with the highest classification success rate), 78 runs led to a fuzzy system whose performance exceeds 96.5%, and of these, 8 runs found systems whose performance exceeds 97.5%. These results are summarised in Fig. 4. Table 2 presents the average performance obtained by the genetic algorithm over all 120 evolutionary runs, divided according to the three experimental categories discussed above.

4.2 Evolving the fuzzy approach

We next describe three of our top-performance systems, which serve to exemplify the solutions found by our evolutionary approach. The first system, delineated in Fig. 5, consists of three rules (note that the ***else*** condition is not counted as an active rule). Taking into account all three criteria of performance, classification rate, number of rules per system, and average number of variables per rule, this system can be considered the top one over all 120 evolutionary runs. It obtains 98.7% correct classification rate over the benign cases, 97.07% correct classification rate over the malignant cases and an overall classification rate (*i.e.* over the entire database) of 97.8%.

However, a thorough test of this three-rule system revealed that the second rule (as shown in Fig. 5) is never actually used; in other words it never fires, *i.e.* is triggered by none of he input cases. Thus, it can be eliminated altogether from the rule base, resulting in a two-rule system (also reducing the average number of variables per rule from 4.7 to 4).

Can the genetic algorithm automatically discover a two-rule system, without resource to any such post-processing? This is indeed the case one such solution is illustrated in Fig. 6. It obtains 97.3% correct classification rate over the *benign* cases, 97.49% correct classification rate over the *malignant* cases and an overall classification rate of 97.36%. Finally Fig.7 delineates the best one-rule system found through our evolutionary approach. It obtains an overall classification rate of 97.07%.

4.3 Diagnostic Confidence

Up until now we have been using the evolved fuzzy systems to ultimately produce a binary classification value, *benign* or *malignant*, with no finer gradations. Referring to Fig.1 we note that the diagnostic systems comprises in fact two subsystems. The first subsystems consists of the (evolved) fuzzy system, which, upon presentation of an input (in our case, a WBCD database entry) proceeds to produce a continuous appraisal value. This value is then passed along to the second subsystem, the threshold unit, which produces the final binary output (*benign* or *malignant*).

The first subsystem (the fuzzy system) is the one evolved in our approach. The threshold subsystem simply outputs *malignant* if the appraisal value is below a fixed threshold value and outputs *benign* otherwise. The user through knowledge of the problem assigns the threshold value at hand.

The appraisal value can accompany the final output of the diagnostic system, serving as a confidence measure. In general, the appraisal value computed by our evolved fuzzy systems is in the range [2,4]. We choose to place the threshold value at 3, with inferior values classified as *benign* and superior values classified as *malignant*. This demonstrates a prime advantage of fuzzy systems, namely, the ability to output not only a (binary) classification, but also a measure representing the system's confidence in its output. For example the three-rule system of Fig. 5 computes intermediate appraisal values (between say 2.4 and 3.6) for 39 cases; these might thus be considered the cases for which we are somewhat about the output.

CONCLUSIONS

In this study we applied a combined a fuzzy-genetic approach to the Wisconsin breast cancer diagnosis problem. Our evolved systems attain, firstly high classification performance (one of the best shown to date) with the possibility of attributing a confidence measure to the output diagnosis, and second the resulting systems involve a few simple rules, and are therefore interpretable.

This experience suggests that the proposed approach is highly effective where such medical diagnosis problems concerned. Future goal is to develop an approach for automatically producing high-performance interpretable systems for real-worlds diagnosis problems.

REFERENCES

1. Michalewicz, Z., *Genetic Algorithms + Data Structures = Evolution Programs*, Springer-Verlag, 1996.
2. Herrera, F., Lozano, M., & Verdegay, J., Generating fuzzy rules from examples using genetic algorithms, in *Fuzzy Logic and Soft Computing*, Bouchon-Meunier B., et al, Eds., World Scientific, 1995.
3. Merz, C., Murphy, P., *UCI repository of machine learning databases*, http://www.ics.uci.edu/~mlearn/MLRepository.html, 1996.
4. Mangasarian, O., Setiono, R. & Goldberg, W-H., Pattern recognition via linear programming: Theory and application to medical diagnosis, in *Large-scale numerical optimisation*, Coleman, T., & Li, Y., Eds., SIAM ,1990.
5. Quinlan, J., *C4.5. Programs for Machine Learning*, Morgan Kaufmann Publs. 1993.
6. Murthy, S., Kasif, S., Salzberg, S., A system for induction of oblique decision trees, Journal of Artificial Intelligence Research, No. 2, 1994, pp1-32.

7. Kermani, B., White, M. & Nagle, H., *Feature extraction by genetic algorithms for neural networks in breast cancer classification*, Proc. of the 1995 IEEE Engineering in Medicine and Biology Int. Conf., 1995, pp 831-832.

8. Setiono, R., *Extracting rules from pruned neural networks for breast cancer diagnosis*, Artificial Intelligence in Medicine, 1996, pp 37-51.

9. Nauck, D., Klawonn, F. & Kruse, R., *Foundations of neuro-fuzzy systems*, Wiley, 1997.

10. Setiono, R., Liu, H, *Symbolic representation of neural networks*, Computer, Vol. 29, No. 3, 1996, pp 71-77.

11. Taha, I, Ghosh, J., *Evaluation and ordering of rules extracted from feedforward networks*, Proc. of the IEEE Int. Conf. on Neural Networks, 1997, pp 221-226.

12. Pedrycz, W. & Valente de Oliveira, J., *Optimisation of fuzzy models*, IEEE Trans. SMC, Vol. 26, No. 4, 1996, pp 627-636.

Parameter	Values	Bits	Quantity	Total bits
P	[1-8]	3	9	27
d	[1-8]	3	9	27
A	[0-3]	2	$9N_r$	$18N_r$

Table 1: Parameters encoding of an individual's genome. Total genome length is $54+18N_r$, where N_r denotes the number of rules.

Training/test Ratio	Performance			Number of variables
	Training set	Test set	Overall	
100% / 0%	---	---	96.97%	3.32
75% / 25%	97.00%	96.02%	96.76%	3.46
50% / 50%	97.71%	94.73%	96.23%	3.41

Table 2: Summary of results of 120 evolutionary runs, divided according to the three experimental categories.

Figure 1: Proposed diagnosis system

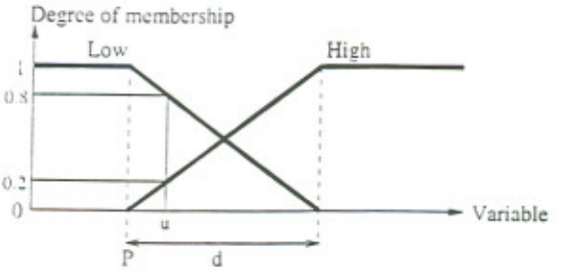

Figure 2: Example of a fuzzy variable with two possible fuzzy values labeled Low and high, and orthogonal membership functions, plotted above as degree of membership versus input values. P and d define the start point and the length of membership function edges, respectively. The orthogonality condition means that the sum of all membership functions at any point is one.

P_1	D_1	P_2	d_2	P_3	d_3	P_4	d_4	P_5	d_5	P_6	d_6	P_7	d_7	P_8	d_8	P_9	d_9
4	3	1	5	2	7	1	1	1	6	3	7	4	6	7	1	1	5

	A_1^1	A_2^1	A_3^1	A_4^1	A_5^1	A_6^1	A_7^1	A_8^1	A_9^1
	0	1	3	3	2	3	1	3	1

(a)

Database

	v_1	v_2	v_3	v_4	v_5	v_6	v_7	v_8	v_9
P	4	1	2	1	1	3	4	7	1
d	3	5	7	1	6	7	6	1	5

Rule Base

Rule 1: **if** (v_2 **is** *Low*) **and** (v_5 **is** *High*) **and** (v_7 **is** *Low*) **and** (v_9 **is** *Low*)
then (*output* **is** *benign*)

Default: **else** (*output* **is** *malignant*)

(b)

igure 3: Example of a genome for a single-rule system. (a) Genome encoding. The first 18 positions encode the parameters P and d for the nine variables v_j. The rest encode the membership function applicable for the nine antecedents of each rule. (b) Interpretation. Database and rule base of the single-rule system encoded by (a).

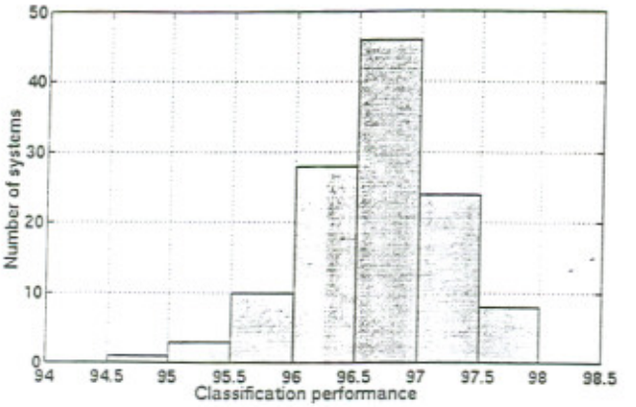

Figure 4: Summary of results of 120 evolutionary runs. The histogram depicts the number of systems exhibiting a given performance level at the end of the evolutionary run. The performance considered is that of the best individual of the run, measured as the overall percentage of correctly classified cases over the entire database.

Database

	v_1	v_2	v_3	v_4	v_5	v_6	v_7	v_8	v_9
P	3	5	2	2	8	1	4	5	4
d	5	2	1	2	4	7	3	5	2

Rule Base

Rule 1: **if** (v_3 **is** *Low*) **and** (v_7 **is** *Low*) **and** (v_8 **is** *Low*) **and** (v_9 **is** *Low*) **then** (*output* **is** *benign*)

Rule 2: **if** (v_1 **is** *Low*) **and** (v_2 **is** *Low*) **and** (v_3 **is** *High*) **and** (v_4 **is** *Low*) **and** (v_5 **is** *High*) **and** (v_9 **is** *Low*) **then** (*output* **is** *benign*)

Rule 3: **if** (v_1 **is** *Low*) **and** (v_4 **is** *Low*) **and** (v_6 **is** *Low*) **and** (v_8 **is** *Low*) **then** (*output* **is** *benign*)

Default: **else** (*output* **is** *malignant*)

Figure 5: The best evolved fuzzy diagnostic system with three rules. It exhibits an overall classification rate of 97.8% and an average of 4.7 variables per rule.

Database

	v_1	v_2	v_3	v_4	v_5	v_6	v_7	v_8	v_9
P	1	1	1		6	2		3	
d	5	3	2		7	4		1	

Rule Base

Rule 1: **if** (v_1 **is** *Low*) **and** (v_3 **is** *Low*) **then** (*output* **is** *benign*)

Rule 2: **if** (v_2 **is** *Low*) **and** (v_5 **is** *Low*) **and** (v_6 **is** *Low*) **and** (v_8 **is** *Low*) **then** (*output* **is** *benign*)

Default: **else** (*output* **is** *malignant*)

Figure 6: The best evolved fuzzy diagnostic system with two rules. It exhibits an overall classification rate of 97.36% and an average of 3 variables per rule.

Database

	v_1	v_2	v_3	v_4	v_5	v_6	v_7	v_8	v_9
P	4	4				2		2	
d	3	1				5		7	

Rule Base

Rule 1: **if** (v_1 **is** *Low*) **and** (v_2 **is** *Low*) **and** (v_6 **is** *Low*) **and** (v_8 **is** *Low*) **then** (*output* **is** *benign*)

Default: **else** (*output* **is** *malignant*)

Figure 7: The best evolved fuzzy diagnostic system with one rule. It exhibits an overall classification rate of 97.07% and a rule with 4 variables.