Independent and Simultaneous Evolution of Fuzzy Sleep Classifiers by Genetic Algorithms

Cristina Mota
LaSEEB-ISR-IST
Av. Rovisco Pais, 1, TN 6.21
1049-001 Lisboa Codex
Portugal

Heitor Ferreira
LaSEEB-ISR-IST
Av. Rovisco Pais, 1, TN 6.21
1049-001 Lisboa Codex
Portugal

Agostinho Rosa
LaSEEB-ISR-IST
Av. Rovisco Pais, 1, TN 6.21
1049-001 Lisboa Codex
Portugal

Abstract

This paper describes two alternative approaches to the automatic inference of a fuzzy classification system applied to computerised sleep staging. Both approaches use genetic algorithms to evolve a fuzzy classifier per sleep stage. In the first case, each stage classifier is independently evolved while in the second case, the classifiers are evolved simultaneously. Satisfactory results where obtained for the individual stage classifiers (76% to 97%), but the global performance of the classification system decreased significantly. No significant differences between the two evolution methods were observed. Possible improvements are suggested.

1 INTRODUCTION

According to the most widely used sleep classification criteria, proposed in 1968 by Rechtschaffen and Kales [1], sleep is a dynamic process which can be divided into 7 stages: wake, NREM1, NREM2, NREM3, NREM4 and paradoxical sleep or REM.

In the clinical context, sleep staging is a routine process that involves the analysis of about 2000 pages of poligraphic recordings. In this light, the demand for an automatic classification tool becomes a necessity.

Since the seventies, several efforts have been taken to automate the sleep staging process. The agreement rates attained between manual and automatic classification ranged from 75% to 89% [8]. These results are quite satisfactory taking into account the inherent subjectivity of the classification criteria. This subjectivity is patent in the inter-individual and inter-laboratorial agreement rates, which range between 67% and 88% [8].

In [7] and [9], two automatic sleep staging systems based in fuzzy logic are described. Although these systems can be helpful tools in the clinic environment, their configuration is a complex process involving the fine-tuning of several interdependent variables. In this article we propose the use of genetic algorithms for the automatic configuration of a fuzzy logic classification system. The classification system is composed by four independent classifiers - one per sleep stage or group of sleep stages - and a module responsible for integrating their outputs.

Two alternative evolution techniques are explored. In the first approach each classifier is evolved independently. Since the global performance of the integrated classifiers is significantly lower than the individual performances of the stage classifiers, a second approach is tried where all classifiers are inferred simultaneously.

The results obtained for the two techniques are discussed and their performance is compared.

2 SYSTEM DESIGN

The classification system architecture is depicted in Figure 1.

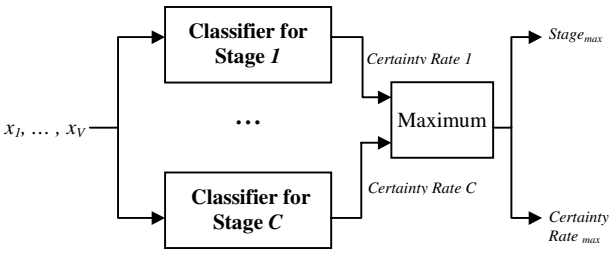


Figure 1: Automatic Classification System

It is composed by 4 fuzzy classifiers (C=4), each one being responsible for the following stages:

Classifier$_1$ - Wake;

Classifier$_2$ - NREM1 and REM;

Classifier$_3$ - NREM2;

Classifier$_4$ - NREM3 and NREM4;

The input for the classification system is the feature vector $x_1, \ldots, x_V$ (V=12) extracted from a segment of EEG recording. The feature vector is presented to each classifier, each one returning a certainty rate. This value is assumed to reflect the classifier's confidence on the recognition of an instance of the represented class.

A set of certainty rates is thus obtained and the end result of the classification system will be the maximum value in this set and an index identifying the classifier which produced it.

If more than one stage classifier outputs the same maximum value, the global result is labelled as indecision

2.1 CLASSIFIER STRUCTURE

Each classifier is structured as depicted in Figure 2.

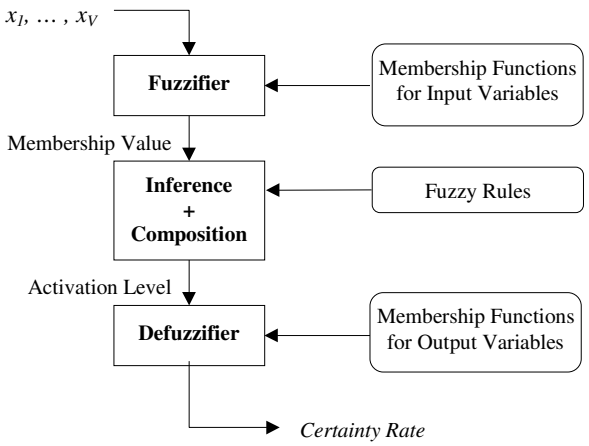


Figure 2: Classifier Structure

The input variables (one per feature) and the output variable (the certainty rate) are characterised by, at most, 3 membership functions. Each variable ranges between 0 and 1 and has the pre-defined shapes depicted in Figure 3.

The following constraints where imposed to the centre and width of each member function, where *MF* is the number of membership functions per variable:

- the width of each membership function ranges in the following interval :

[0;max-min] = [0;1]

- The centre c_i (i=1,...,*MF*) ranges in the interval:

$$\left[(i-1)\times\frac{\max-\min}{MF};i\times\frac{\max-\min}{MF}\right]=\left[\frac{i-1}{MF};\frac{i}{MF}\right]$$

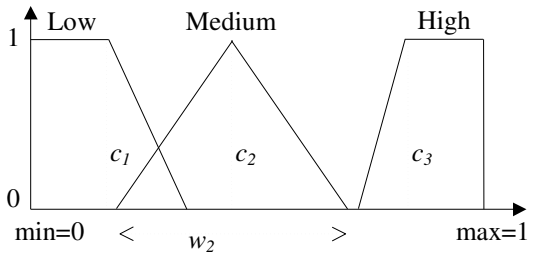


Figure 3: Membership Functions

A rule is structured as follows:

IF <expression> THEN <expression>

where <expression> is

- A literal with the structure: *variable IS membership-function*, possibly modified by the negation operator NOT.
- A sequence of literals connected by the operator AND.

A variable appears only once in a rule. The following rule, for example, is not well formed:

IF ($\mathbf{v_1}$ IS m_1) AND ($\mathbf{v_1}$ IS m_2) THEN (v_{output} IS m_1)

The product and sum methods were used in the inference and composition steps; defuzzification used the "average of the maximum" to calculate the crisp value.

It was established that the rule set would contain, at most, 10 rules.

3 FEATURE EXTRACTION

To test the classification system, EEG recording for several individual where obtained from the C4-A1 electrode placement, according to the 10-20 international system.

The selection of relevant features for this work was based on the visual classification criteria proposed by Rechtschaffen and Kales [1]. For simplicity, only the tonic activity was considered: a total of 12 features are extracted to characterise the activity in the frequency bands normally used in the clinical context:

Delta	[0 Hz, 4 Hz[
Teta	[4 Hz, 8 Hz[
Alfa	[8 Hz, 13 Hz[
Beta	[13 Hz, 30 Hz]
Sigma	[12 Hz, 16 Hz]

The frequency band of relevant activity ([0Hz, 30Hz]), enclosing all the above specified bands, is also considered.

The spectral analysis is based on the power spectra density estimation by the Welch Periodogram algorithm.

Each band is characterised by the following features:

- Medium band power - the medium power in the band
- Band bias - the power bias in the band, computed as for the total bias. It characterises the power distribution in the band

Accordingly, the band of relevant activity is characterised by the following features:

- Total power - the estimated power in the band [0 Hz, 30Hz]
- Total bias - characterises the power distribution. It is the medium frequency in the band [0 Hz, 30 Hz] scaled to the interval [0, 1]

To minimise inter-individual variations, the medium band power is normalised by the individual's medium power obtained in the REM stage.

This normalisation implies the existence of a manual classification of the EEG recording or the implementation of a REM stage detector (possibly using EOG recordings). Since the goal of this work is to validate the classification system according to manually classified data, the implementation of a REM detector was not attempted.

4 INDEPENDENT EVOLUTION

In this approach the parameters of each classifier are inferred independently. This means that in a certain population there are only individuals responsible for the identification of a certain stage. From the evolution of each stage is chosen the best individual, using as criterion the success rate for a test set with all patterns different from the training set. These individuals will parameterise the global classification system.

4.1 CODING

Each individual is formed by two genomes.

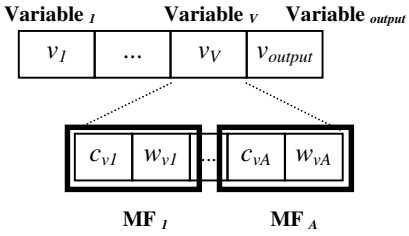
4.1.1 Genome representing the parameters of membership functions

Defines the centre and width of each membership function for each fuzzy variable.

Supposing there are V+1 variables (V input variables and one output variable) and MF membership functions, this genome has a length equal to L_{mf}:

$$L_{mf} = (V + 1) * (2 * MF)$$

and it is represented by the following schema:



c_{ij} represents the centre of membership function j of variable i;

w_{ij} represents the width of membership function j of variable i.

If the width of any membership function is 0 this membership function will be removed from the set of membership functions of the variable to which it refers.

Each centre or width is represented by an 8-bit word, creating a set of $2^8 = 256$ possible values. Therefore, this genome can represent T_{mf} different membership functions:

$$T_{mf} = 256^{L_{mf}}$$

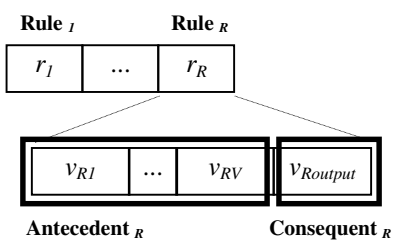
4.1.2 Genome representing the rules

Defines the rule set of the classifier. Each rule is composed by $V+1$ genes that represent each variable: the first V genes form the antecedent; the last one forms the consequent. Each gene has $2*MF+1$ possible alleles that correspond to MF membership functions, plus MF negated membership functions plus one additional value that controls the presence or absence of the correspondent variable in the rule.

Supposing that the maximum number of rules is R then the size of this genome is L_{rules}:

$$L_{rules} = (V+1) * R$$

and is represented by the following schema:



If the membership function (negated or not) in gene v_{ij} has a width equal to 0, the variable associated with it will no

longer be considered in the antecedent. If the same happens with the membership function associated with the variable of the consequent the correspondent rule will no longer belong to the rule set of the classifier.

This genome can represent T_{rules} possible rules:

$$T_{rules} = (2 * MF + 1)^{L_{rules}}$$

4.2 GENETIC OPERATORS

4.2.1 Initialisation

The population is initialised randomly.

4.2.2 Reproduction

The individuals for reproduction are chosen by the roulette wheel method that uses the fitness function described in section 4.2.4.

4.2.3 Crossover

Given two individuals the crossover is done by applying the uniform crossover operator to each genome that composes the individuals (see Figure 4).

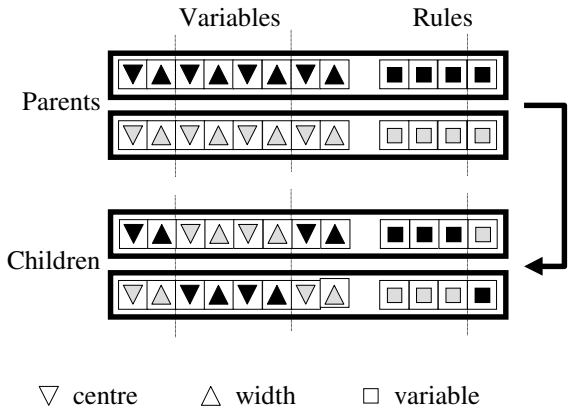


Figure 4: Crossover between two classifiers

The crossover points are defined randomly for each crossover. However the crossover points of the variables do not separate a centre of a membership function from its width.

4.2.4 Mutation

The mutation of the genome describing the variables consists of choosing randomly a centre or a width in the neighbourhood of the current value (Figure 5).

Within the genome that describes the rules the normal operator is applied (Figure 5).

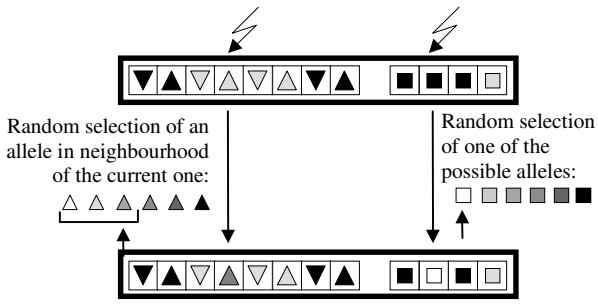


Figure 5: Mutation of a classifier

4.3 FITNESS FUNCTION

Supposing that a classifier is responsible for the identification of a stage s then its fitness f is given by the following expression:

$$f = \frac{1}{\overline{D} + k}$$

Where,

$$\overline{D} = \frac{\sum_{i=1}^{P} D(d(x_i) - o(x_i))}{P}$$

$$D(d(x_i) - o(x_i)) = \left| d(x_i) - o(x_i) \right|$$

$$d(x_i) = \begin{cases} 1, C(x_i) = s \\ 0, C(x_i) \neq s \end{cases}$$

$\overline{D}$ represents the average distance to the goal.

$C(x_i)$ is the correct classification of pattern x_i , $o(x_i)$ the output of the classifier which reflects the certainty rate and P the total number of patterns in the training set.

Constant k defines the limits for the fitness value. The results presented below were obtained with k=0.001.

5 SIMULTANEOUS EVOLUTION

In simultaneous evolution each individual represents a classification system formed by several fuzzy classifiers, one for each stage.

5.1 CODING

In this approach the individual is formed by C genomes coded as described in the previous section (see 4.1), meaning that each classifier is parameterised,

independently from the others, by a variable set and a rule set (Figure 6).

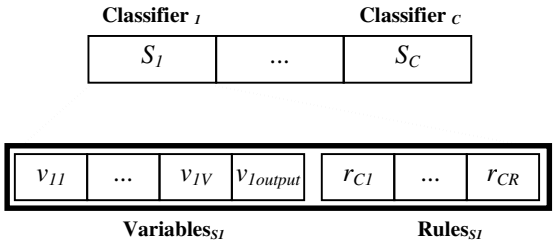


Figure 6: Coding of a classification system

5.2 GENETIC OPERATORS

5.2.1 Initialisation

The population is initialised randomly.

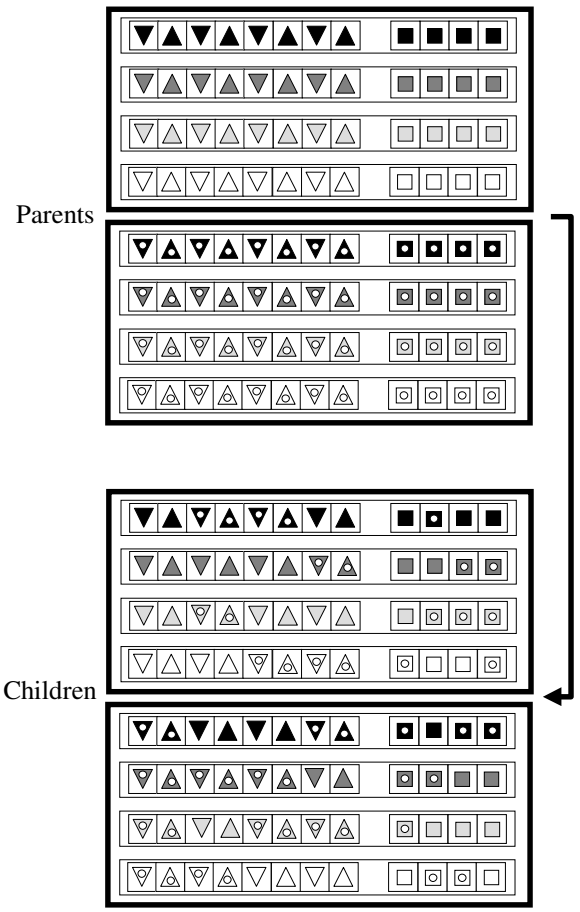


Figure 7: Crossover between two classification systems

5.2.2 Reproduction

The individuals for reproduction are chosen by the roulette wheel method that uses the fitness function described in section 5.3.

5.2.3 Crossover

Each classifier in a system is crossed with the correspondent classifier in the other system. This means that the crossover is done applying the crossover operator described in section 4.2 using the classifiers responsible for the identification of the same stage as parents (see Figure 7).

Each colour corresponds to a different type of classifier; the circle corresponds to a different system. So the crossover is done between genomes that have the same colour but one has the circle and other does not.

5.2.4 Mutation

This mutation consists of applying the mutation operator described above in section 4.2 to each classifier.

5.3 FITNESS FUNCTION

Two fitness functions were implemented: f_1 is similar to the one described in section 4.2.4 but considers the Euclidean distance between two patterns; f_2 is based upon the success rate of the system.

$$f_1 = \frac{1}{\overline{D} + k} \qquad f_2 = \frac{1}{1 - \frac{\sum_{i=1}^{P} d_j}{P} + k}$$

Where,

$$\overline{D} = \frac{\sum_{i=1}^{P} D(d(x_i) - o(x_i))}{P}$$

$$D(d(x_i) - o(x_i)) = \left\| d(x_i) - o(x_i) \right\|$$

$d(x_i)$ is a binary vector that has only one of the positions set to 1 (this position identifies the classifier that must produce the highest certainty rate), $o(x_i)$ is a vector formed by the certainty values produced by each classifier and d_j is the value of the position j of the vector $d(x_i)$, where j is the index of the classifier that produced the highest certainty rate. When more than one classifier produces the highest certainty rate $d_j=0$ (it counts as a fault).

6 RESULTS

Several preliminary tests were performed to determine the evolution parameters that best fitted this particular problem. Unless stated otherwise, all the results presented in this paper were obtained with the parameters found with these tests, which are detailed in Table 1.

Table 1: Evolution Parameters

	Independent Evolution	Simultaneous Evolution
Population	50	50
Elitism	Yes (1 individual)	Yes (1 individual)
Crossover	90%	90%
Mutation	0,5%	0,3%
Generations	500	500

Elitism was implemented by maintaining the best individual from one generation to another.

The experiments were made using each patient recording in particular (R1, R2 and R3) and all recordings mixed together (R4).

In individual evolutions the training and test sets are formed by equally distributed positive and negative examples: the number of positive examples, of the stage to identify, is equal to the number of negative ones.

The patterns for simultaneous evolutions are classified in four stages (one for each classifier).

The training and test sets are disjointed.

In the graphics, we used the following notation:

- Fitness Evolution: the grey dashed line represents the fitness of the best individual, the black line represents the average fitness of the population and the grey solid line represents the fitness of the worst individual;
- Error Evolution: the grey solid line represents the average distance to the goal, the black line represents the success rate and the grey dashed line represents the indecision rate for the best individual.

6.1 INDIVIDUAL EVOLUTION

As it may be verified in Figure 8, in individual evolutions the fitness of the worst individual suffers great oscillations and does not increase significantly along evolution.

It may also be observed that after the first 50 generations, although the average distance decreases, the success rate does not increase significantly (see Figure 9)

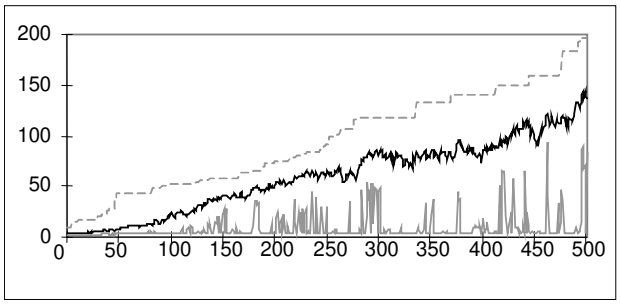


Figure 8: Fitness Evolution of an NRem1/Rem classifier

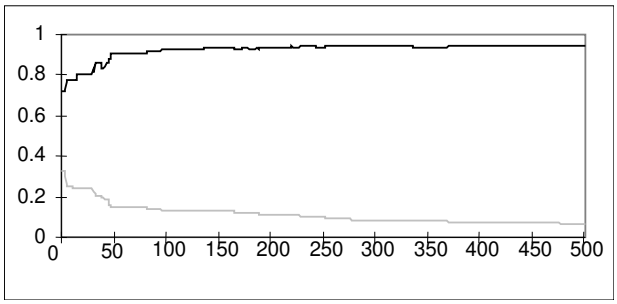


Figure 9: Error Evolution of an NRem1/Rem classifier

As table Table 2 demonstrates, each classifier has excellent results with the training set, but these results decrease significantly with the test.

Table 2: Success rates, in %, for classifiers trained individually with *R3/R4*

	Training	Test	Simultaneous Test
Wake	92/86	84/79	88/68
NRem1 Rem	97/88	91/84	74/75
NRem2	90/81	88/76	56/65
NRem3 NRem4	95/90	91/88	92/45
			77/63

This decrease is more significant in the simultaneous test (when the classifiers are combined in the classification system) where the global success rate has very low values - 63% for R1, R2 and R4 and 77% for R3. For all simultaneous tests the indecision rate was below 3,5%.

The system also displays difficulty in identifying the Wake and NRem2 stages. This behaviour further damages the performance of the classification system.

The output value of each classifier is normally restricted to two values only - 0 and a value in the neighbourhood of 1 – thus the interpretation of this value as a certainty value is inappropriate. The decision by the highest output value is favourable to the classifiers that systematically produce the highest output values, which may not necessarily correspond to the classifiers with best performance.

6.2 SIMULTANEOUS EVOLUTION

The most striking difference introduced by the simultaneous evolution is the fact that the worst individual's fitness improved with time (Figure 10 and Figure 11).

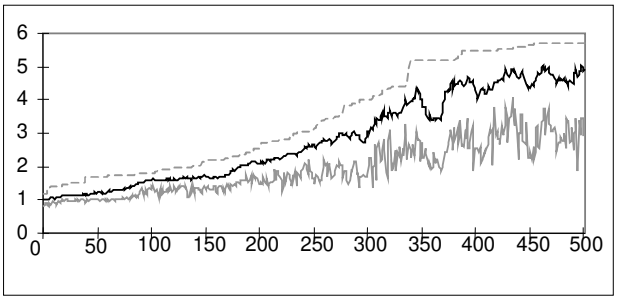


Figure 10: Fitness evolution using f_1

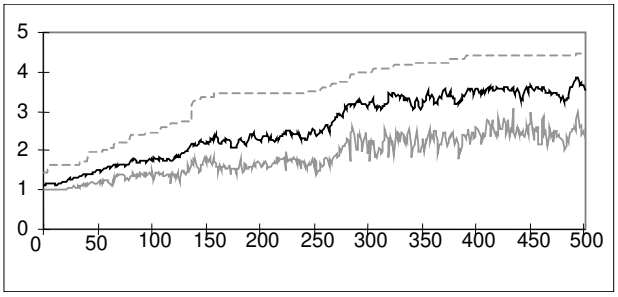


Figure 11: Fitness evolution using f_2

It was observed that although the population fitness continues rising above the 400th generation, the success rate does not improve significantly (Figure 10 and

Figure 11). Typically, after 500 more generations, the success rate did not improve more than 6%.

The indecision rates are typically higher when using f_1 than when using f_2. Besides, the success rate evolves slower with f_1 than with f_2.

As for performance with the test sets, the success rates decreased about 5% to 9% relative to the training sets. The difference between the classifiers evolved with f_1 and f_2 increases (Table 3): classifiers based in f_2 display lower indecision rates than f_1 based classifiers, which present indecision values ranging form 8% to 44%.

Table 3: Success rates, in %, for R1, R2 and R3

	f_1		f_2	
	Training	**Test**	**Training**	**Test**
R1	61,6	56,7	71,2	66,7
R2	62,7	54,3	70,6	61,6
R3	76,0	67,3	79,6	69,6

Although the evolutions performed with data from 3 individuals presented lower training success rates, the observed decrease from training to test sets was not so evident.

The results obtained for the R4 sets are similar to those obtained for the independent evolution of stage classifiers.

Similarly to what was observed for the independent evolution, the stage classifiers present dissimilar performances and this is reflected by the global system performance (Table 4).

Table 4: Error matrix for the classifier trained with *R4* (training and test)

Training/ Test	**Wake**	**NRem1 Rem**	**NRem2**	**NRem3 NRem4**
Wake	76 / 80	18 / 11	4 / 5	1 / 3
NRem1 Rem	14 / 19	78 / 68	5 / 10	2 / 2
NRem2	4 / 7	13 / 20	68 / 43	15 / 30
NRem3 NRem4	4 / 5	2 / 0	10 / 15	83 / 81

7 FUTURE WORK

The uneven performance of the stage classifiers, together with the fact that the classifiers frequently output the values 0 and 1, instead of a continuous spectrum of values, suggest an integration strategy other than the one based in the maximum output value. A possible alternative would be to rank the classifiers according to their test performance. A pattern rejected by a classified (output lower than a given threshold) would be presented to a lower rank classifier and so forth. A pattern accepted by a classifier (output above the threshold) would be affected with the respective stage and not presented to the

lower rank classifiers. Indecision would occur if the lowest rank classifier rejected a pattern.

Another methodology to be explored is to combine the two evolution techniques. First the individual classifiers would be independently evolved and then the best individuals in each population would be combined to generate a new population of global classifiers that would be further evolved.

8 CONCLUSIONS

Although the performance of the global classification system, either evolved independently or simultaneously, is not very encouraging, the individual performance of the stage classifiers is satisfactory.

Relatively to the simultaneous classifier evolution, it was concluded that a higher system performance is obtained (better test success rates and lower indecision rates) when the evolution is carried with the fitness function based in the success rate (f_2).

It was observed that some of the stage classifiers performed better than others. This indicates that further investigation should be carried concerning the selection of relevant features. Besides, it is possible that some of the used features are not relevant for this classification problem, and the excessive number of features can degrade the performance of the evolution algorithm.

Acknowledgments

This work is partially funded by PRAXIS /2/2.1/SAU/1392/95.

Bibliography

[1] Rechtschaffen, A., and Kales, A., *A manual of Standard Terminology Technics and Scoring System for Sleep Stages of Human Subjects,* Washington, NIH Pub, 1968.

[2] D. Dubois and H. Prade, *Fuzzy Sets and Systems: Theory and Applications.* New York: Academic Press, 1980.

[3] H. Zimmermann, *Fuzzy Set Theory - and its Applications,* 2nd ed. Boston: Kluwer, 1990

[4] A. Homaifer e E. McCormick, "Simultaneous Design of Membership Functions and Rule Sets for Fuzzy Controllers Using Genetic Algorithms", IEEE Trans. on Fuzzy Systems, vol. 3, no. 2, pp129-139, May 1995.

[5] J. H. Holland, *Adaptation in Natural and Artificial Systems*. Ann Arbor, MI: University of Michigan, 1975.

[6] D. E. Goldberg, *Genetic Algorithms in Search, Optimization and Machine Learning.* Addison-Wesley, 1889.

[7] N. Martins, F. Freire, C. Matos, C. Rosa and T. Paiva, "Automatic Sleep Scoring of EEG Recordings, Using Fuzzy Logic", in Proceedings of the 7th International IMEKO TC-13 Conference on Measurement in Clinical Medicine, pp.170-173, September1995.

[8] N. Schaltenbrand, R. Lengelle, M. Toussaint, R. Luthringer, G. Carelli, A. Jacqmin, E. Lainey, A.Muzer and J.P. Macher, Sleep Stage Scoring Using the Neural Network Model: Comparison Between Visual and Automatic Analysis in Normal Subjects and Patients, Sleep no. 19, pp26-35, 1996.

[9] G. Bonfiglioli, C. Rosa, C. Narduzzi, *Classificazione Automatica del Sonno Basata Sulla Logica Fuzzy,* 1996.

[10] L. A. Zadeh, "Outline of a New Approach to the Analysis of Complex Systems and Decision Processes", IEEE Transactions on Systems, Man and Cybernetics, Vol. SMC-3, no. 1, pp. 28-44, 1973.

[11] J. S. Barlow, *The Electroencephalogram - Its Patterns and Origins,* Cambridge, Massachussests: MIT Press, 1993.

[12] T. Paiva*, Sono: Aspectos Clínicos e Funcionais*, Teste de Doutoramento, Faculdade de Lisboa, 1991.

[13] B. Kosko*, Neural Network and Fuzzy Systems*, Prentice Hall, New Jersey, 1992.

[14] H. Ishibuchi, K. Nozaki, H.Tanaka, "Selecting Fuzzy If-Then Rules for Classification using Genetic Algorithms", IEEE Transactions on Fuzzy Systems, vol. 3, no. 3, pp. 260-270, August 1995.