AN EVOLUTIONARY APPROACH TO SIMULATE COGNITIVE FEEDBACK LEARNING IN MEDICAL DOMAIN

H.S. LOPES,
Centro Federal de Educação Tecnológica do Paraná / CPGEI,
Av. 7 de setembro, 3165 - 80230-901, Curitiba (PR) - Brazil

M.S. COUTINHO,
Universidade Federal de Santa Catarina / GPEB,
P.O. Box 476, 88040-970, Florianópolis (SC) - Brazil

W.C. de LIMA,
Universidade Estadual de Santa Catarina,
P.O. Box 1228, 88080-350, Florianópolis (SC) - Brazil

Cognitive feedback is a technique used for qualitative learning that has proven to be useful to train medical students. In this work we report the application of genetic algorithms to simulate this technique, using a knowledge-based system as the learner, in the domain of coronary artery disease diagnosis. The prototypical description of the disease employs fuzzy variables, as well as crisp ones. To evaluate the performance of the system, a similarity-triggered inference method is used over a diagnosed case-base. Results presented showing the efficiency of this approach, lead us to believe that this paradigm is useful for a wide range of applications.

keywords: genetic algotithms, surface analogy, learning, medical diagnosis.

1 The Learning Paradigm

1.1 Quantitative versus qualitative learning

Most developed knowledge-based systems are strongly domain specific: small changes in the application domain usually reflect large changes in the system's knowledge-base. An intensive human intervention to deal with these changes has to take place, in order to properly re-adjust the knowledge-base. This is what Holland[1] has called "brittleness". Actually, "brittleness" basically reflects the incapacity of some current computational systems of generating new knowledge or adapting current knowledge, by means of its own acquired experience. In contrast, one of the most startling characteristic of human intelligence is its ability to learn by experience and evolve its own knowledge, when interacting with the environment. Therefore, the learning capability turns out to be a necessary feature for high MIQ (Machine Intelligence Quotient) systems.

This is specially true for systems designed to be used in noisy or changing environments. In fact, there are several different, but somewhat superimposed, definitions of learning to span many kinds of processes. Here, the meaning of learning is understood in the sense stated by Carbonell[2]:

> "A system (biological or mechanical) is said that to learn if it can modify its behavior after a set of experiences such that it can perform a task either more accurately or more efficiently than before, or perform a new task beyond its previous capabilities."

Based on this definition, two forms of learning can be devised:

- Quantitative learning: infer new knowledge, not initially belonging to the knowledge-base.
- Qualitative learning: improve performance, refining the current knowledge-base.

The first could be exemplified as a rule-based system that applies deductive inference to generate new facts that will be further integrated into the knowledge-base. A classical example of the second form is a neural network, which, starting from scratch (no knowledge), learns patterns using some gradient descent technique, achieving better performance the more trained it is. Qualitative learning is generally achieved by active experimentation rather than passive observation. It is related to the correction of observed deviations from a desired behavior by means of repetitive training. In this type of learning, feedback plays a key role because it drives the learning process.

An important point about learning, although not much clear, is its relationship to a measurement procedure. It is always necessary to measure performance so as to evaluate whether or not an improvement has occurred and in what extent. This is true for both human and machine learning.

1.2 Cognitive feedback as a qualitative learning paradigm

There has been a close link between Cognitive Science and Artificial Intelligence since the beginning. Machine learning techniques have been inspired in human learning paradigms. In the other hand, computer simulations have helped researchers understand how humans learn.

Cognitive feedback is a technique that has been applied to human learning and has been object of study by some researchers (see Balzer et al[3], for instance) since it was first proposed by Todd and Hammond[4]. In this technique, a judgment is considered as a function that takes into account the multiple features that describes the problem. Each feature is related to the judgment (and the actual outcome) through a weight. By means of the judgment of a series of cases, one infers weights related to the features. Correct weights, relating features to the actual outcome, are then presented to the learner to provide a feedback. Although the practical implementation of this technique allows di®erent variations, the feedback provided to the learner is the main factor of judgment accuracy improvement, as reported by Hammond [5].

The use of this technique as a pedagogical instrument in medical domain has been successfully reported, as for instance Poses et al [6] and Wigton et al [7]. The e®ect on medical students was a qualitative improvement of their diagnostic performance of cases in the speci¯ed domain.

The main objective of this work is to simulate a kind of cognitive feedback technique, using a knowledge-based system as learner, which performance it is expected to improve.

2 Framework for diagnosis

2.1 Motivation and knowledge representation

Evidences have shown that expert clinicians perform, although not always conscientiously, diagnoses based on analogy with memorized prototypes of diseases[8,9,10]. When analyzing a case, the more similar it is to a disease prototype, the more plausible will be the possibility of the patient to have the disease.

Based on this idea, a framework for diagnosis in medical domain was developed [11]. In this system, knowledge is represented in a frame structure as a disease prototype. Each slot of the frame comprises a relevant information (attribute) for the ¯nal diagnosis. All attributes have a value which represents the typical ¯nding for a positive diagnosis. A weight is attached to each attribute of the frame, representing the discriminatory power (or \diagnosticity") of current information for ¯nal diagnosis. Therefore, a disease prototype is represented by a set of m triplets attribute-value-weight ($D = \{a_i; v_i; w_i\}^m$). A clinical case is described by a set of attribute-value pairs of the same nature of the prototype and, for a given patient, $P = \{a_i; p_i\}^m$. The set of pairs value-weight is used in the computation of the similarity level between a patient's data (P) and the disease prototype (D), using syntactic analogy [a] and

[a] This kind of analogy is also known as surface analogy and is related to the attributes that are syntactically identical in two objects, not accounting for contextual or structural factors.

a geometric model, as in equation 1, for the Euclidean distance.

$$\Delta(P;D) = \sqrt{\sum_{i=1} \left(w_i \cdot \frac{|p_i - v_i|}{\mathrm{range}(v_i)} \right)^2} \tag{1}$$

The set $\{a_i; v_i; w_i\}$ for each prototype is elicited from the domain expert, in a three-step procedure: first the relevant attributes, followed by their typical values for that disease, and finally, the relative weights. Representing knowledge using prototypical approach provides a structured representation for classes of diseases. The adequacy and expressivity of this representation is remarkable for using with the similarity-based inference engine.

2.2 Coronary artery disease

Coronay artery disease (CAD) is a common cause of disability and mortality in the adult population. It is associated with risk factors such as smoking, high cholesterol, arterial hypertension, obesity, diabetes and physical inactivity. The main symptom in CAD is chest pain, but unfortunately this symptom is a common one for a group of other frequent diseases as well. Diagnosing CAD is not an easy task, even for the experienced physician. A thorough medical examination includes a detailed clinical history and physical examination, complemented by non-invasive laboratory investigations (exercise test, cholesterol and glucose levels, echocardiography and nuclear imaging). Coronary arteriography, an invasive and definitive procedure, is indicated only when a surgical intervention (coronary artery bypass grafting) is planned or the informations from clinical assessment and non-invasive procedures are not sufficient to conclude a diagnosis. In the real-world, physicians must live with uncertainty which is counteracted by increasing expertise and experience.

2.3 CAD prototype with fuzzy variables

Several factors are accounted for the diagnosis of CAD, including symptoms, observed signals and complementary test results. Depending on the physician's experience and available resources, some factors may be considered, others not. In this work, it was used a database of 303 CAD cases [b] that was previously employed in classification studies. This database contains 13 attributes which description and respective values characterizing a typical CAD case are summarized in table 1. All attributes are normalized in the $[0;1]$ interval, and

[b] In fact, only 297 cases were used, the 6 other instances were eliminated due to missing attributes.

Table 1: Description of the coronary artery disease (CAD) prototype (? see text).

#	variable	typical CAD	type
1	age	> 45	fuzzy?
2	sex	male	binary
3	chest pain	typical angina	numeric
4	resting systolic blood pressure	> 140 mmHg	fuzzy?
5	serum cholesterol	> 240 mg/dl	fuzzy?
6	fasting blood sugar	> 120 mg/dl	binary
7	resting ECG	T inversion or ST elevation or ST depression > 0:05 mV	numeric
8	maximum heart rate achieved	< 130 bps	fuzzy?
9	exercise induced angina	yes	binary
10	induced ST depression	> 1:0 mm	fuzzy?
11	slope of ST in peak exercise	down sloping	numeric
12	colored vessels in fluoroscopy	3	numeric
13	thallium results	fixed defect	numeric

further combined by the analog reasoner of the inference engine (see figure 4). Details of the implementation of the analog reasoner are beyond the scope of this work and are reported elsewhere[12].

In medicine, more than in physical sciences, it is commonplace uncertain reasoning. Several factors contribute to this: first, clinical findings may largely vary from case to case. Second, there is inherent fuzzyness in some clinical variables. Third, a diagnosis may have to be done basing on incomplete, inaccurate or even inconsistent data. Last, all patient's data is under the subjectiveness of the physician's interpretation.

In clinical diagnosis, like other real-world applications, frequently we have to deal with concepts with ill-defined boundaries. Humans have a natural ability to assign grades of membership to complex concepts without being explicitly aware of this. The physician expert whom the values of table 1 were elicited from, could express better his knowledge about the diagnosis using intervals of uncertainty for some attributes. Those considered qualitatively were attributes #1, #4, #5, #8 and #10. Some mathematical approaches have been employed to deal with uncertainties in medical domain (see Hughes[13], for instance), but a more natural one for this case is to use fuzzy hedges[14]. Considering those attributes as linguistic variables, a fuzzy set described by a membership function can be constructed. To model membership functions, a parameterized approach devised by Dombi[15] was used, as follows:

$$\mu(x) = \begin{cases} \dfrac{(1-v)^{\lambda-1}(x-a)^{\lambda}}{(1-v)^{\lambda-1}(x-a)^{\lambda}+v^{\lambda-1}(b-x)^{\lambda}} & \text{ascending part} \\[2ex] \dfrac{(1-v)^{\lambda-1}(b-x)^{\lambda}}{(1-v)^{\lambda-1}(b-x)^{\lambda}+v^{\lambda-1}(x-a)^{\lambda}} & \text{descending part} \end{cases} \qquad (2)$$

Where v is the inflection point of the S-shaped curve, λ is a measure of sharpness of the curve, and a and b represent lower and upper bound of support such that $a < b$, and $\mu(a) = 0$ and $\mu(b) = 1$, for an ascending curve; or $\mu(a) = 1$ and $\mu(b) = 0$ for a descending curve. The domain expert has arbitrarily assigned values to parameters a and b, while v and λ were respectively set to $(b - a)/2$ and 1 (see table 2).

Table 2: Values of the Dombi model for the CAD prototype. (A = ascending curve, D = descending curve).

variable	parameter				type
	a	b	λ	v	
age	30	50	1	40	A
resting blood pressure	120	160	1	140	A
serum cholesterol	180	220	1	200	A
maximum heart rate achieved	110	150	1	130	D
induced ST depression	0.5	1.5	1	1.0	A

In this approach, it is considered separately the monotonical increasing and decreasing parts of a membership function. Thus, for a S-shape curve, either the upper or the lower part of equation 2 is used to represent the uncertainty interval, and for the rest of support, $\mu(x)$ is either 1 or 0. Actually, $\{a; b; \lambda; v\}$ represent a family of curves, as shown in figures 1 and 2 for the interval $[0; 1]$, and some values of parameters λ and v.

3 Implementation

3.1 Genetic algorithms overview

Genetic algorithms [16;17] (GAs) are a search technique inspired in Darwinian natural selection principle, and have been applied to several classes of optimization problems.

Parameters of a problem, representing a candidate solution, are coded into a string, usually of fixed length λ and using the binary character set $\{0; 1\}$. Strings are individuals that compose a population of size n. Each individual

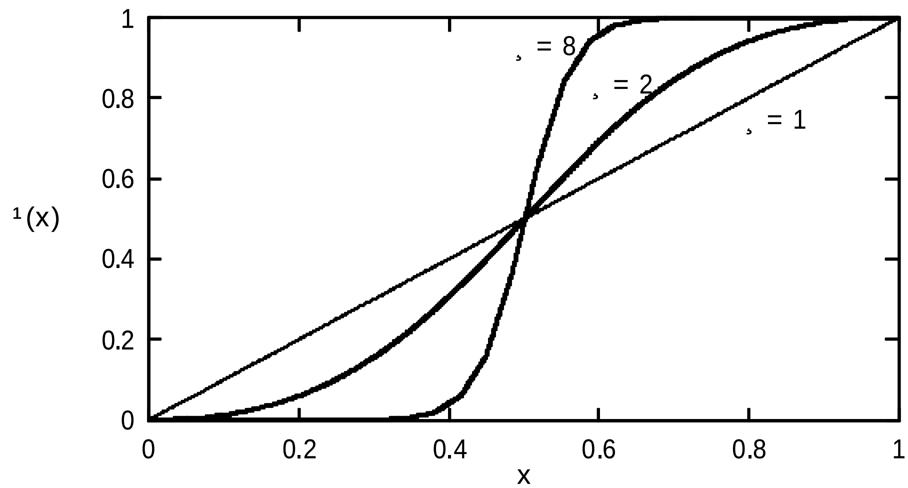

Figure 1: Ascending curves of membership functions representing the uncertainty interval $[0;1]$ for three different values of λ.

$a_i^t = \{0;1\}^\lambda$ is evaluated according its ability to provide an useful solution, by means of an objective (or fitness) function $f(a_i^t)$.

GAs work iteratively, where each cycle t corresponds to a new generation of individuals, starting with a ramdomly choosen population $P(0) = (a_1^0; \ldots; a_n^0)$. Every new generation $P(t+1) = (a_1^{t+1}; \ldots; a_n^{t+1})$ is a result of a reproduction process to which was submitted the former population $P(t) = (a_1^t; \ldots; a_n^t)$. The reproduction process undergoes a selective pressure, resembling the natural selection principle of biological evolution. This implies that individuals will be selected as parents according to a probability proportional to its fitness, as in equation 3. By means of genetic operators, such as crossover and mutation, selected parents produce offsprings. A population of individuals is then evolved throughout generations. As generations go by, the population is expected to converge to individuals of high fitness level, and thus, good solutions.

$$prob(a_i^t) = \frac{f(a_i^t)}{\sum_{k=1}^{n} f(a_k^t)} \qquad (3)$$

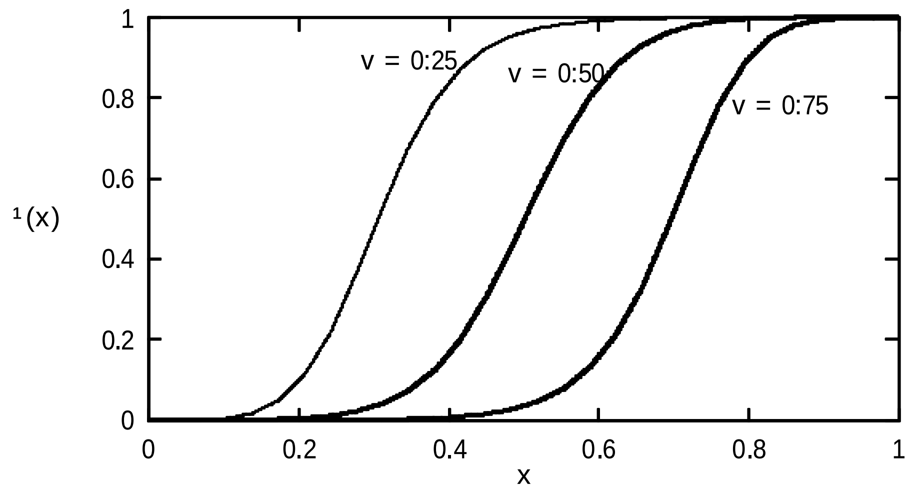

Figure 2: Effect of varying v (inflection point) on an ascending curve of membership function in the interval $[0; 1]$.

GAs are included in a class of techniques known as Evolutionary Algorithms, which are distinguished by their parallel investigation of the search space, by manipulating a population of possible solutions simultaneously. Being a general-purpose search method, the applicability of GAs is narrowed when problem-specific knowledge is incorporated to the system aiming performance improvement. For a given problem, the performance of the algorithm is severely influenced by a number of parameters, which have to be carefully choosen. Besides, two other factors, how an individual is encoded and the objective function, play an important role in the overall functioning, and will be discussed in the next sections.

3.2 Encoding and running parameters

The most usual representation for GAs is the binary encoding, although others have been used for specific purposes. The set of weights of CAD prototype was encoded into a 52-bit binary string representing an individual, thus, each real-valued weight was mapped using 4 bits length. The choice of this representation has no biological inspiration, but relies on the fact that this quantization step has acceptable accuracy for the problem. Figure 3 shows how weights are encoded into the chromosome. This representation yields a search space of $2^{52} \approx 10^{16}$ points, a reasonably large space for conventional

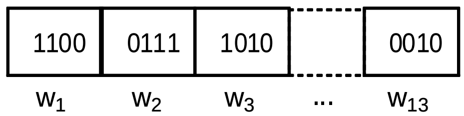

Figure 3: Chromosome representation.

methods. A GA-based method for searching the global optimum is then specially suited, considering not only the high dimensionality of the problem, but also its multimodal nature. In fact, it is not always expected to ¯nd the global optimum, but in many cases a near optimal solution within some percentage of the optimal is satisfactory. This approach was called \satis¯cing" by Simon[18], in order to indicate that the candidate solution satis¯es minimal requirements and su±ces to reasonably solve the problem.

A GAs module is responsible for applying genetic operators in order to constitute a population of individuals that will be evolved throughout generations. The probability of reproduction of each individual is proportional to its performance, following equation 3. Each individual of each generation is evaluated by an objective function (see section 3.3), and its ¯tness is computed. At the end of each generation, the best individual is decoded to update weights of CAD prototype. Figure 4 shows a simple block diagram of the system.

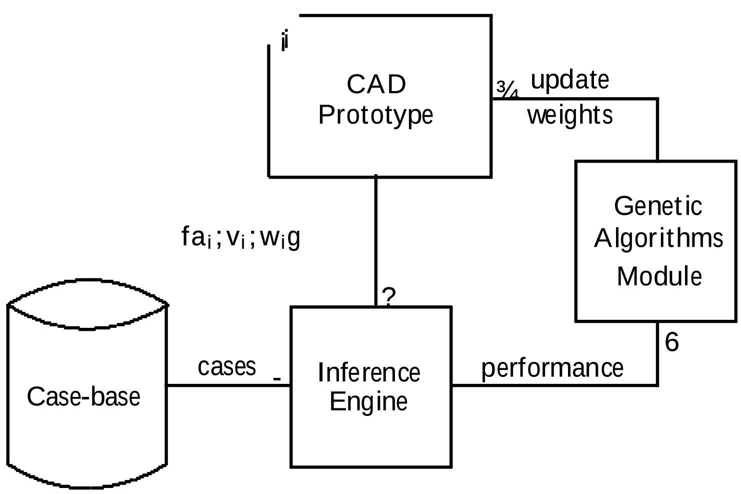

Figure 4: Block diagram of the system for cognitive feedback learning.

The set of weights was initially set to a random number between 0:00 and 1:00, representing absolute ignorance about relative relevance of each information. When applying GAs to optimize the prototype's set of weights, a kind of qualitative learning is implemented, simulating the cognitive feedback technique, previously mentioned. GAs can be run using a large range of control parameters. Changing parameters will deeply a®ect the performance of the algorithm. Inadequate settings may induce premature convergence as well as \slow ¯nishing", the opposite problem. Several trials with di®erent sets of parameters were done, and the best results was observed when the following parameters were used:

- initial population: randomly generated
- crossover probability: 1.0
- mutation probability: 0.005
- type of selection: proportional, non-elitist
- population size: 100 individuals
- generation gap: 1 (non-overlapping populations)
- length of each individual: 52 bits, ¯xed
- maximum number of generations: 500

3.3 Objective function

The task of classifying unknown cases of CAD may involve two types of errors: a healthy patient being misclassi¯ed as sick (false-positive), and a sick patient misclassi¯ed as healthy (false-negative), while a correct classi¯cation should lead respectively to true-positive and true-negative. In medical domain, two parameters are used as performance measures of a classi¯er test: sensitivity (S) and speci¯city (E), de¯ned as follows:

$$S = \frac{tp}{tp + fn} \tag{4}$$

$$E = \frac{tn}{tn + fp} \tag{5}$$

where $tp; tn; fp; fn$ stands respectively for true-positive, true-negative, false-positive and false-negative. Sensitivity measures the fraction of patients having CAD that will be correctly detected by the system, and speci¯city measures the fraction of healthy patients who will be correctly identified as having no disease.

If sensitivity is increased, patients having no disease may be misclassified as having disease (increases false-positive rate). On the other hand, if specificity is increased, patients having CAD may be misclassified as healthy (increases true-positive rate). Thus, sensitivity and specificity are mutually constrained.

A graphics plotting S versus $(1 - E)$ results in a concave curve known as "receiver-operating characteristic" curve (ROC). The optimum point for S and E depends on the relative costs associated with the misclassification of diseased patients as healthy, versus healthy patients as diseased. Here, costs are generally related to health factors rather than economic factors. For the purposes of this work, the objective is to find the point in the ROC curve where both S and E are maximum. At this point, the maximum number of correct classifications will be attained, not biasing either S or E. Therefore, a trade-off between these two performance measures has to be considered, leading to a kind of multicriteria optimization.

The objective function is the natural selection criterion, and should return a single value representing the fitness level of the individual. Several combinations of the sensitivity and specificity measures were considered, and better results were obtained using simply the product between sensitivity and specificity, as shown in equation 6. "Better results" have a meaning in accordance to the previous considerations about relative costs.

$$fitness = S.E \tag{6}$$

This function considers both measures simultaneously and, as they are normalized in the $[0, 1]$ range, fitness values are also normalized in this range. To compute sensitivity and specificity for an individual (set of weights), a program routine loops over a data base of 297 previously diagnosed cases, applying the methodology described by Lopes et al [12]. Figure 5 shows the core pseudocode of the evaluation procedure.

4 Discussion and conclusions

Comparing initial diagnostic performance of CAD diagnostic system (with all weights randomly choosen), to the best-fitted set of weights (of all generations), it was reported a significant improvement in performance. An accuracy of 81.5% of correct diagnoses was achieved by this method, slightly outperforming other methods previously reported in literature, using the same database [19,20,21,22]. Actually this seems to be around the upper limit for clas-

```
== evaluate k-th individual: fai; wig
FOR n = 1 TO 297 DO
    == compute diagnosis using case(n) and individual(k)
    outcome = diagnosis(case(n), individual(k))
    == compare outcome with correct diagnosis and update counters
    compute tp, fp, tn, fn
ENDDO
== compute ¯tness for individual k
S = f(tp,fp,tn,fn)
E = f(tp,fp,tn,fn)
fitness(k) = f(S,E)
```

Figure 5: Evaluation procedure of an individual (set of weights).

si¯cation accuracy using this database. The learning curve shown in ¯gure 6 displays two points where the best accuracy was attained.

Considering the noisy feature of the database, an improvement in the robustness of the system was observed when fuzzy variables were used instead of crisp ones. Furthermore, the use of fuzzy concepts has proved to be useful also in the elicitation task phase. We believe that for real-world problems, specially medical diagnosis, fuzzy thinking is an essential feature for computational systems aiming to achieve human-like performances.

We point out training the system as the major drawback, because setting GA control parameters has shown not to be trivial. Furthermore, as the objective function has revealed highly multimodal, additional care must be taken to assure proper convergence. As in other real-world problems, it is di±cult to establish an optimum balance between selective pressure, which accelerate convergence, and population diversity, which assures proper search space exploration. This conclusion is in accordance with researchers that consider setting GA control parameters an ad hoc problem [23;24]. Notwithstanding, results presented demonstrate the feasibility of GA-based learning to improve the diagnostic ability of the system.

Although a threshold was used to discriminate diagnosed cases into two distinct classes (healthy or not healthy), the system is able to provide a contin-

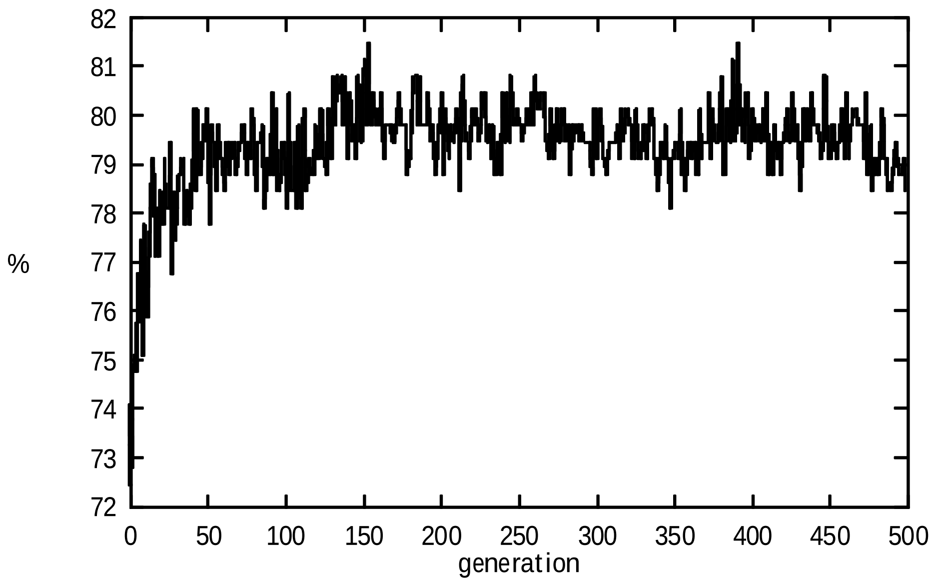

Figure 6: Learning curve showing the percentage of correct diagnoses as a function of the generation epoch.

uous classi¯cation, allowing a graceful response as input data quality degrades. This is possible due to the similarity-triggered analogical reasoner and the prototypical representation of the knowledge-base. The outcome provided by the system may be roughly understood as a membership degree to a multidimensional fuzzy set of a typical CAD patient.

The core system has proved to be an e±cient paradigm to deal with complex diagnostic tasks, and the objective of simulating the cognitive feedback technique was sucessfully achieved. The credible performance displayed by the system, encourages both further improvements and its application to other domains.

Acknowledgements

This work was partially supported by a CAPES-CNPq grant to H.S. Lopes. We would like to thank Robert Detrano (V.A. Medical Center, Long Beach and Cleveland Clinic Foundation) for providing the Cleveland heart disease database. This database is available from University of California at Irvine repository of machine learning databases. Also, authors are grateful to Prof. Jorge Barreto (UFSC) for his useful comments and suggestions.

References

1. J. A. Holland, \Escaping brittleness: the possibilities of general purpose learning algorithms applied to parallel rule-based systems," in Machine Learning: an Arti¯cial Intelligence Approach 2, eds. R. S. Michalski, J. G. Carbonell, and T.M. Mitchel, Morgan Kau®mann, Los Altos, (1986), pp. 593{623.
2. J. G. Carbonell, \Derivational analogy: a theory of reconstructive problem solving and expertise acquisition," in Machine Learning: an Arti¯cial Intelligence Approach 2, eds. R. S. Michalski, J. G. Carbonell, and T. M. Mitchel, Morgan Kau®mann, Los Altos, (1986), pp. 371{392.
3. W. K. Balzer, M. E. Doherty, and R. O'Connor, \E®ects of cognitive feedback on performance," Psychol. Bull. 106 (1989) 410{433.
4. F. J. Todd and K. R. Hammond, \Di®erencial e®ects in two multiple-cue probability learning tasks," Behav. Sci. 10 (1965) 429{435.
5. K. R. Hammond, \Computer graphics as an aid to learning," Science 172 (1971) 903{908.
6. R. M. Poses, R. D. Cebul, R. S. Wigton, and M. Collins, \Feedback on simulated cases to improve clinical judgment," Med. Descis. Making 6 (1986) 274.
7. R. S. Wigton, R. M. Poses, M. Collins, and R. D. Cebul, \Teaching old dogs new tricks: using cognitive feedback to improve physicians' diagnostic judgments on simulated cases," Acad. Med. 65 (1990) s5{s6.
8. M. Blois, \Conceptual issues in computer-aided diagnosis and the hierarchical nature of medical knowledge," J. Med. Phylos. 8 (1983) 29{50.
9. H. G. Schmidt, H. P. Boshuizen, and G. R. Norman. \Re°ections on the nature of expertise in medicine," in Deep Models for Medical Knowledge Engineering, ed. E. Keravnou, Elsevier, Amsterdam, (1992), pp. 231{248.
10. U. BÄckenholt and R. Zacks, \Toward a theory of hypothesis generation in diagnostic decision making," Invest. Radiol. 28 (1993) 76{80.
11. H. S. Lopes, M. S. Coutinho, and W. C. Lima, \An application of analogical reasoning to diagnostic tasks," Proc. EXPERSYS'94 (1994) 729{734.
12. H. S. Lopes, M. S. Coutinho, and W. C. Lima. \A model of similarity-triggered analogical reasoning for diagnosis in medicine," Proc. 8^{th} Int. Symp. Syst. Model. Contr. 2 (1995) 13{17.
13. C. Hughes, \The representation of uncertainty in medical expert systems," Med. Inform. 14 (1989) 269{279.
14. L. A. Zadeh, \The concept of a linguistic variable and its application to approximate reasoning," parts 1, 2, 3, Inf. Sci. 8 301{357; 9 (1975)

43{80.

15. J. Dombi, \Membership function as an evaluation," Fuzzy Sets and Syst. 35 (1990) 1{21.
16. J. H. Holland, Adaptation in natural and arti¯cial systems, The University of Michigan Press, Ann Arbor, (1975).
17. D. E. Goldberg, Genetic Algorithms in Search, Optimization and Machine Learning, Addison-Wesley, Reading, (1989).
18. H. A. Simon, The sciences of the arti¯cial, MIT Press, Cambridge, (1969).
19. R. Detrano, A. Janosi, and W. Steinbrunn, \International application of a new probability algorithm for the diagnosis of coronary artery disease," Am. J. Cardiol. 64 (1989) 304{310.
20. D. W. Aha, D. Kibler, and M. K. Albert, \Instance-based learning algorithms," Mach. Learn. 6 (1991) 37{66.
21. J. H. Gennari, P. Langley, and D. Fisher, \Models of incremental concept formation," Artif. Intell. 40 (1989) 11{61.
22. S. Salzberg, \Distance metrics for instance-based learning," in Proc. ISMIS'91 (1991) 399{408.
23. J. J. Grefenstette, \Optimization of control parameters for genetic algorithms," IEEE-SMC 16 (1986) 122{128.
24. J. D. Scha®er, R. A. Caruana, L. J. Eschelman, and R. Das, \A study of control parameters a®ecting the performance of genetic algorithms for function optimization," in Proc. ICGA'89 (1989) 51{60.