

Multipopulation Genetic Programming Applied to Burn Diagnosing.

F. Fernandez de Vega

Dpto. Informática
Univ. de Extremadura
Escuela Politécnica. Cáceres
fcofdez@unex.es

Laura M. Roa

G.I.B. E.S.I
Universidad de Sevilla
Camino de los Descubrimientos
lmroa@esi.us.es

Marco Tomassini

Institute d'Informatique
Univ. de Lausanne
Switzerland
mtomassi@iissun4.unil.ch

J.M. Sanchez

Dpto. Informática
Universidad de Extremadura
Escuela Politécnica. Cáceres
sanperez@unex.es

Abstract- Genetic Programming (GP) has proved useful in optimization problems. The way of representing individuals in this methodology is particularly good when we want to construct decision trees. Decision trees are well suited to representing explicit information and relationships among parameters studied. A set of decision trees could make up a decision support system. In this paper we set out a methodology for developing decision support systems as an aid to medical decision making. Above all, we apply it to diagnosing the evolution of a burn, which is a really difficult task even for specialists. A learning classifier system is developed by means of Multipopulations Genetic Programming (MGP). It uses a set of parameters, obtained by specialist doctors, to predict the evolution of a burn according to its initial stages. The system is first trained with a set of parameters and results of evolutions which have been recorded over a set of clinic cases. Once the system is trained, it is useful for deciding how new cases will probably evolve. Thanks to the use of GP, an explicit expression of the input parameter is provided. This explicit expression takes the form of a decision tree which will be incorporated into software tools that help physicians in their everyday work.

1 Introduction

Making good diagnosis is sometimes only possible when doctors have a strong background of knowledge in a particular medical area. This knowledge has usually been acquired over many years of work. Even specialists find it difficult to make some kinds of diagnoses. It is thus interesting to develop software tools containing large knowledge systems capable of helping doctors in their every day work. These decision systems should contain information directly extracted from clinical cases under the supervision of specialists. The aim is to make these systems as accurate as possible.

In this paper, we are firstly presenting a methodology for developing and constructing decision support systems for medical purposes; and secondly, we are applying this methodology to the problem of burn diagnosis.

When somebody suffers a burn it is necessary to find out its degree, in order to apply the best treatment from the beginning. This classification problem is recognized as

being difficult, because it is not always clear how much tissue damage there is and how it will evolve.

This paper is a part of a larger research project in which we are trying to represent by means of decision trees the knowledge used by specialists when making diagnoses.

The aim is to develop a classifier system capable of analyzing the picture of a burn and automatically detecting its degree.

This investigation has been divided into two different parts: the first sets out a methodology based on Genetic Programming [koza]. Its intention is to collect doctors' medical knowledge. The second part applies this methodology to building a knowledge system containing doctors experience. The system must be capable of predicting how a burn will evolve, allowing users to choose the most suitable treatment. If the methodology presented here is useful in this problem, it can be probably used in other medical fields.

We think evolutionary algorithms have an important advantage over other kinds of machine learning techniques: They have the ability to provide information not only about classification, but also about the route taken to reach the decision.

We are proposing a methodology which works with GP to develop and represent decision trees. These trees are capable of correctly classifying burns using a set of parameters extracted from them

This paper is structured in the following way: The second section sets out a Genetic Programming (GP) technique applied to the construction of decision trees. Section 3 deals with the training method for the system. Section 4 presents our results, and finally section 5 offers our conclusions.

2 Rule Extraction by means of Genetic Programming.

2.1 Classifying by means of decision trees

Making medical decisions is a classification problem, in which the search space is made up of a set of points with n-coordinates. Each coordinate is allocated a value for a given symptom. The aim is to find the category to which a point belongs. This data will give us the diagnosis.

Neural networks can be employed for collecting information and using it in classification problems. Nevertheless, once the network has been trained, the

knowledge stored within is hidden. Thus no new knowledge can be retrieved from the network.

Sometimes we are not only interested in obtaining a classifier but also in retrieving important information from the problem's input parameters. Several classifier systems for medical diagnoses have been designed based on decision trees and decision rules [Pahlm], [Holmes]. They take the parameters and classify the data.

2.2 Decision Trees.

The Decision Trees approach [Quinlan] falls into a range of inductive learning methods. Taking a set of examples the aim is to construct a tree that is able to classify new samples within the search space.

Decision trees usually classify the members of a set as either positive or negative examples. In our research, however, due to the number of possibilities involved in burn diagnosis, we extend the classic approach with multi-class decision making, as in other medical diagnosis research [Kurzynski].

In fact a decision tree is a chain of *if_then_else* construction that can be seen as a computer program (See figure 1)

Each condition is applied to some specific input parameters. The parameters will thus lead us along a branch of the tree, finally reaching a decision, which is the category to which the input data belongs.

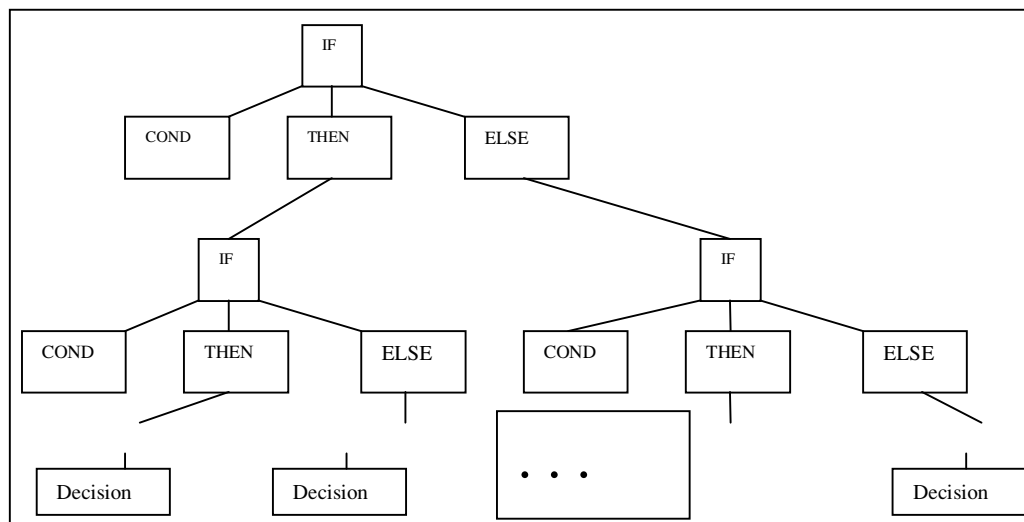


Figure 1: A typical decision tree

We want to generate automatically the decision tree according to a group of clinic cases that make up the training set. This will be done by means of Parallel Genetic Programming, the shape of the decision tree, which actually resembles a program.

2.3 Genetic Algorithms and Genetic Programming.

Genetic Algorithms (GAs) is a kind of optimization method based on natural evolution. Each parameter of a given problem is codified by means of a series of bits. All the bits

make up the chromosome of an individual. Via evolution GAs try to optimize each of the parameters of the problem in hand.

GA have been widely used for solving difficult problems in many fields. We can find research in which GA's are used for classification problems in medical tasks [Kurzynski], [Azorman], [Matsofoulos].

Genetic programming is the automated learning of computer programs. GP is inspired in the theory of evolution and assembles variable length programs structures from basic units called functions –i.e. instructions - and terminals –i.e their inputs- Some genetic operators are used to promote diversity: crossovers between two individuals, mutation and finally selection of the fittest. Applying these operations to a number of generations, we expect to find individuals, in other words programs, capable of solving a previously established problem.

We obtain the set of functions and the set of terminals by selecting those that are helpful in the solving of the problem.

Bearing in mind that in GP each individual adopts the shape of a tree, a decision tree can be considered to be an individual. If we apply GP to finding an specific decision tree, at the end of the process the best individual takes the form of that decision tree, i.e. the genotype of the individual presents the chain of decisions which are necessary for a diagnosis to be made.

When using GP for decision tree extraction, the function set will only be composed of the *If_then_else* instruction, and some logical operators (See section 3). Thus any tree can be constructed. The terminal set will contain the parameters we have decided to study in each burn. Thus an individual could consequently feature shapes like those depicted in figure 2.

2.4 Multipopulations Genetic Programming (MGP).

New GP models have been proposed and studied with the aim of improving results and reducing the time required to obtain them. Multipopulation Eas and specifically MGP

allow us to work simultaneously with several population, saving computing time.

We have been working with a kind of multipopulation model: coarse-grained models.

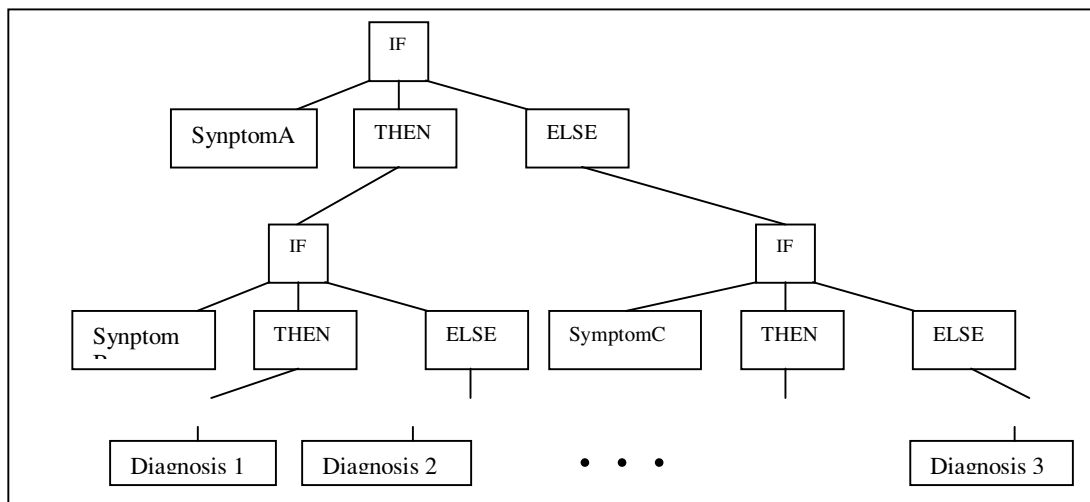


Figure 2: In our problem an individual is a decision Tree.

In coarse grained models, also called island parallel models, the population of individuals is divided into several autonomous subpopulations, usually called demes. Demes are allowed to exchange individuals at a certain rate (the migration rate) according to a usually predefined communication topology. The main reason for using this approach is its ability to avoid premature convergence by injecting new individuals, which promote diversity, while at the same time exploring different portions of the search space within each deme.

Within each deme, a standard sequential evolutionary algorithm is executed between migration phases. Several migration policies have been described [Fernandez et al 1], [Oussaidène et al]. The most common one replaces the worst k individuals of a deme with the same number of individuals coming from other different populations, usually copies of the best individuals.

The topology of communication is also important. The most common topologies are ring structures, 2-d and 3-d meshes. Hypercubes and random graphs have also been employed. Here we introduce a dynamic topology, described in the following section, in which the exchange pattern changes during the run.

Not only in GP but also in GA employing several populations has proved to be useful in obtaining good results [Cantú-Paz, Goldberg], [Oussaidène et al], [Fernandez et al 1]. In particular, several papers have dealt with the number of populations and individuals that must be used for optimizing results [Cantú-Paz, Goldberg], [Lin, Punch][Oussaidène et al].

2.5 Our Multipopulation Genetic Programming Model.

A software tool was designed for working with MGP [Fernandez et al 2]. This tool uses the client/server (Figure 3) model for establishing the communication topology. Every population acts as a client process, sending its best individual to the server every time an internal measuring-point is reached. The server is in charge of deciding where each received individual must travel, so establishing the communication topology.

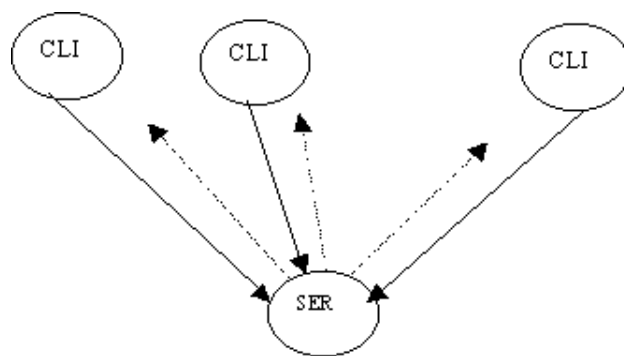


Figure 3: Client/Server Model. Clients are Subpopulations.

Our previous research showed us that random topology, in which the destination of an individual is always stochastically generated (see figure 4), gives good results in a wide range of problems [Fernández et al 1]. We thus decided to use such a topology.

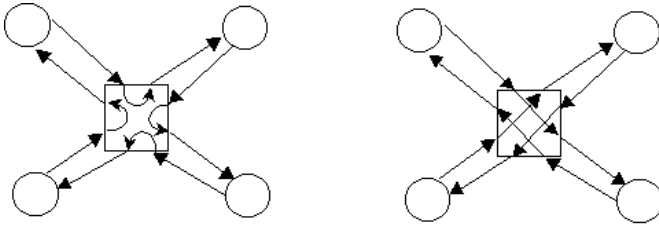


Figure 4: Random Topology: Each time the communication topology can randomly change.

2.6 The parameter set

Following specialist doctors' advice, we decided to develop the first version of the decision tree taking into account just three different parameters from each picture:

- Color (C): Several possible values: White, Black, Yellow, Red, Pink and combinations.
- Dryness (D): Two possibilities: True or False.
- Ampoule (A): Two possibilities: Absent or Not.

Our aim in this first stage was to develop a Knowledge System based on a decision tree which could classify burns. We asked doctors for a set of clinic cases where the parameters described and the diagnoses of the burn were recorded. Four kind of diagnosis are possible: 1st degree, surface 2nd degree, deep 2nd degree and 3rd degree.

Although studying a wider set of parameters is more useful, this simple model confirm the validity of the methodology employed.

At the moment, it is not very important to have a tool to classify 100% of the cases. The aim is to demonstrate how easily doctor's knowledge can automatically be captured by means of GP and decision trees.

3 EXPERIMENTAL RESULTS

31 clinical cases were used to train the system. These samples were recorded in Virgen del Rocío Hospital's Burns Unit in Seville, Spain. The specialists were provided with a form to be completed for each treated burn. The burn was allocated its corresponding parameter and the result of its evolution after a couple of weeks. This table of values was given to the algorithm in order for it to train. Of course, the ultimate aim for the research is to obtain automatically the classification of burns by means of image processing. Nevertheless, we must first obtain the decision tree. This is done by providing the system with the set of clinical cases.

Tables 1 and 2 show the Function set and the Terminal set used in solving the problem at hand.

Table 1: Function Set with number of descendant

Function	If_then else	>	=	<	Nand	Nor
Number of	3	2	2	2	2	2
Descendants						

Table 2: Terminal Set

Terminal	C	D	A
Number of possible values	15	2	2

We ran MGP with 5 populations of 2500 individuals each, with a period of migration of 10 generations. We waited for 60 generations before taking the results.

In the end, the best decision tree we obtained was able to classify correctly 27 out of 31 cases. This was due to the presence of several cases with the same parameters but with different evolution values. It was consequently impossible to categorize these cases accurately. Thanks to the use of this methodology, we are able to representing medical knowledge by means of decision trees. These trees will be integrated into larger software tools capable of analysing and predicting how burns will evolve, thus helping in deciding the best treatment. Nevertheless, bearing in mind that it is impossible to diagnosing correctly all the samples, the results are very encouraging. Figure 5 shows the decision tree.

```
GP: (IF (IF (IF A D DEEP-2ND-DEGREE)
(NAND (> C_YELLOW C) C) SURFACE-2ND-
DEGREE) (IF D 3RD-DEGREE SURFACE-2ND-
DEGREE) (IF (IF D (IF S (> C_PINK C) (>
C_PINK C)) (= C_PINK C)) DEEP-2ND-DEGREE
SURFACE-2ND-DEGREE)) )
```

Fig: 5 The Decision Tree Obtained

4 CONCLUSIONS.

GP is a powerful tool for obtaining decision trees. Not only have we developed a classifier system, but also generated explicit information that shows relationships among parameters. This is very useful in understanding how each of them influences the problem.

Thanks to MGP, automatically representing medical information is an easy task. No effort is necessary to decide which is the best way of modeling data. Providing a set of examples is enough for evolution to retrieve relationships among parameters and build a decision tree. This tree can later be inserted in a software tool which allows us to make diagnoses. We have shown the usefulness of this methodology in burn diagnosing, with three parameters. The

model can be completed with as many parameters as necessary, and the evolution will retrieve the new set of relationships among parameters by means of a new decision tree.

We are at present linking this part of our research to image processing, in order to obtain an autonomous burn classification system.

References

- [Pahlm et al 90] Pahlm, D; Case, D; Howard, G; Pope, J; Haisty, WK. "Decision rules for the ECG diagnosis of Inferior Myocardial Infarction". *Comput Biomed Res.* 1990, Aug; 23(4): 332-45
- [Holmes] Holmes, John H; "Discovering risk of disease with a learning classifier system". *Proceedings of the seventh international conference on Genetic Algorithms.* Pp 426-433.
- [Kurzynski] Kurzynski, Marek; "The application of unified and combined recognition decision rules to the multistage diagnosis problem". *Proceedings of the 20th Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, vol 20, No 3, 1998.
- [Azorman et al] Azorman, Milan; Kokol, Peter; Molan Stiglic, Milojka; Gregoric, Aljoz; "Are decision trees way around some statistic methods?". *Proceedings of the 20th Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, vol 20, No 3, 1998.
- [Podgorelec et al] Podgorelec, Vili; Kokol, Peter; "Evolutionary Construction of Medical Decision Trees.
- [Quinlan] Quinlan J.R. "Decision trees and instance-based classifiers", in *CRC Handbook of Computer Science and Engineering*, ed. A.B. Tucker, CRC Press, 1996.
- [Cantú-Paz, Goldberg] Cantú Paz and David Goldberg: "Predicting Speedups of Ideal Bounding Cases of Parallel Genetic Algorithms". *Proceedings of the Seventh International Conference on Genetic Algorithms.* Morgan Kaufmann.
- [Oussadéne et al] M. Oussaidéne, B. Chopard, O. V. Pictet, , M. Tomassini: "Parallel Genetic Programming and its application to trading model induction". *Parallel Computing*, 23, pp. 1183-1198, 1997.
- [Lin, Punch] S-C Lin, W.F. Punch and E.D. Goodman, "Coarse-grain Genetic Algorithms, Categorization and New Approaches" *Sixth IEEE Parallel and Distributed Processing* Oct 94, pg. 28-37.
- [Fernandez et al 1] F. Fernandez, M. Tomassini, J.M. Sanchez "Solving the ant and the Even Parity Problems by means of Parallel Genetic Programming". *Gecco 99 Late Breaking papers.* Pp.88-92.
- [Fernandez et al 2] Francisco Fernández, Juan M. Sánchez, Marco Tomassini and Juan A. Gómez: " A parallel Genetic Programming Tool Based on PVM". Jack Dongarra, Emilio Luque, Tomás Margalef (Eds) *Recent Advances in Parallel Virtual Machine and Message Passing Interface.* Springer. 1999.
- [Ercal et al] F. Ercal, M. Moganti, W. V. Stoecker and R. H. Moss. *IEEE Transactions on Medical Imaging*, Vol, 12, No 13. SEP 1993. 624-627
- [Umbaugh et al] Umbaugh S.E., Moss R.H. Stoecker W.V.: *An Automatic Color Segmentation Algorithm with Application to Identification to Skin Tumor Borders.* *Comp. Med. Imaging and Graph.* 13(3): 227-235. May-Jun 1992.
- [Matsofoulos et al] George K. Matsofoulos, Nicolas A. Mouravliasky, Konstantinos K. Pelabasis, Konstantina S. Nikita. "Automatic Retinal Image Registration Scheme using Global optimization Techniques". *IEEE Transactions on Information Technology in Biomedicine*, Vol. 3, No. 1, March 1999.