

# Learning of Fuzzy Cognitive Maps for simulation and knowledge discovery

Gonzalo Napoles<sup>1</sup>, Isel Grau<sup>1</sup>, Ricardo Pérez-García<sup>1</sup>, Rafael Bello<sup>1</sup>

<sup>1</sup>Universidad Central “Marta Abreu” de Las Villas, Santa Clara, Cuba  
{gnapoles, igrau, rpgarcia, rbellop}@uclv.edu.cu

## Abstract

In recent years Fuzzy Cognitive Maps (FCM) has become a useful Soft Computing technique for modeling and simulation. They are connectionist and recurrent structures involving concepts describing the system behavior, and causal connections. This paper describes two abstract models based on Swarm Intelligence for learning parameters characterizing FCM, which is a central issue on this field. At the end, we obtain accurate maps, allowing the simulation of the system and also the extraction of relevant knowledge associated with underlying patterns.

**Keywords:** FCM, modeling, simulation, learning, knowledge discovery.

## 1. Introduction

The Cognitive Mapping theory was originally proposed by R. Axelrod [1] who focused on policy domain studies, that is, for modeling social scientific knowledge. A simple Cognitive Map is composed by two components: nodes and connections [1]. Nodes or concepts represent variables describing the system; while links among concepts are used to regulate the causality and take values either -1 or +1.

The sign associated to each link have a relevant meaning for the system interpretation. For example, if the sign of a relationship is positive (+1), then an increase or decrease of the first variable causes the effect variable to change in the same direction (e.g. an increase in the first variable causes increase of the effect variable); whereas in the case of negative relationships (-1), the change of the effect variable is in the opposite direction (e.g. an increase in the first variable causes a decrease of the effect variable).

However, the capabilities of Cognitive Maps for representing the causality of the problem domain are quite limited since it is complicated to define the semantic of a whole system by only expressing a sign on relationships, without a specific degree of causality in causal relations.

For addressing this weakness, B. Kosko introduced Fuzzy Cognitive Maps [2] thus enhancing Cognitive Maps. More specifically, instead of only using a sign, each cause-effect relationship is associated with a weight in range [0, 1]. It will determine the degree of causality between two nodes (variables). This fact allows implementing relevant knowledge concerning with the quanti-

tative strength of each relationship, which now can be described by a fuzzy term, such as: weak, medium, strong or very strong [3].

Recently FCM theory has gained a lot of attention among researchers. For example, in [4] the authors propose a FCM for studying travel behavior in modern societies allowing policy-makers better understanding of these issues. As another example, [5] describes a model based on FCM for analyzing the dynamics of HIV protease protein; enabling to discover relevant knowledge when a mutation takes place. Also, FCM theory has been widely used in other applications fields including: engineering, risk analysis, business, decision making tasks, management, system control, medicine, game theory, and also telecommunications [6], etc.

Furthermore, its strength is really associated with its adaptability to a given domain, comprehensible structure and control, flexibility of system design, and its capability of abstract representation and fuzzy reasoning [7]. Though, tasks related with FCM modeling and design could be quite complex or tedious for humans due to the highly number of sensible aspects that should be considered (e.g. estimation of the causal weights matrix). In fact, the formulation of new learning algorithms for tuning modeling parameters is still an open problem for researchers.

This work describes two abstract learning models based on Swarm Intelligence for optimizing parameters in a FCM. The first model is oriented to the estimation of the causality among concepts, while the second one is designed to determine central nodes in the map. Consequently, fully optimized maps are obtained, allowing better knowledge representation and reasoning. In addition, we present two study cases: one concerning to the protein modeling field, where resulting maps are used for simulation and knowledge discovery, and another related with travel behavior analysis, where central variables describing the problem are determined.

The rest of the paper is organized as follows. In the next Section we provide a mathematical background of FCM. In the Section 3 we describe the learning model for estimating the causality degree among system variables; while in the Section 4 the second learning algorithm is presented, which is capable to determine central nodes on the graph. Afterwards, Sections 5 and 6 illustrate both algorithms through the study cases. To close, some remarks about simulations and the knowledge discovery process are discussed.

## 2. Fuzzy Cognitive Maps theory

Fuzzy Cognitive Maps (FCM) are a type of Recurrent Neural Network used for knowledge representation and reasoning [2]. It combines the connectionist features of Artificial Neural Networks, with some aspects of Fuzzy Logic in their representation scheme. In general terms, a FCM is capable to describe the behavior of a system by using a collection  $n$  of concepts or nodes and weighted relations.

Concepts in the connected graph represent entities, states, objects or variables describing the problem domain, whereas weighted arcs denote the cause-effect relation among system variables. Each concept  $C_i$  is characterized by a fuzzy value in range  $[0,1]$  denoting a degree of activation of the concept at the time  $t$ . In other words, the activation level of a concept is related with the presence degree of a specific system feature at the time  $t$ .

Also weighted arcs have associated a fuzzy value in range  $[0,1]$  or  $[-1,1]$ . Thus, the weights of the system's interrelations reflect the degree of causal influence between two concepts and they are usually assigned linguistically by experts [8].

From the mathematical point of view, a FCM may be characterized by a state vector  $C_{1xn}$  denoting the activation value of concepts, and a weight matrix  $W_{n \times n}$  grouping the causality degree among variables. The inference process in a FCM consists on computing the current state vector  $C_{1xn}$  through time, for a fixed initial condition. It means that the concepts' value will be influenced by connected concepts with the appropriate weights, as the equation (1) shows, therefore showing the effect on varying the value of a concept over the whole map. In next formula,  $S$  denotes a transformation function to keep the concept's activation value in range [9].

$$C_i^{(t+1)} = S\left(\sum_{j=1}^n C_j^{(t)} w_{ij}\right), i \neq j \quad (1)$$

The inference process is repeated till a maximum number of times  $T$  is reached, or an attractor equilibrium state (called *hidden pattern*) is observed. When an attractor point is found the state vector  $C$  won't suffer significant changes in subsequent iterations. Though, in some cases it is possible that the inference mechanism falls into a *chaotic state*. It means that the FCM model will produce different state vectors for successive cycles.

### 2.1. Aggregation of FCM

Usually, a FCM denotes the knowledge extracted from an expert in a given area. But it is possible to have better consistency in the drawing of a FCM if more than one expert is used in the system modeling [2]. The aggregation process of different expert's criteria into a single structure allows improving the reliability of the final model by combining incomplete or conflict opinions from different knowledge sources. This process makes the final map less

susceptible to the undesirable effects of ignorance and erroneous beliefs.

The possibility of aggregating several FCM into a single knowledge structure is an important advantage over other well-known knowledge based approaches such as Bayesian Networks or Petri Nets; since is not obvious for a non-expert users how to combine in a single structure different Bayesian Networks or Petri Nets describing the same system [5].

The number of experts in the FCM aggregation is not limited, in fact, the more experts are involved in the modeling process, the more reliability the final system has [3]. On the other hand, the Kolmogorov strong law of large numbers ensures that, with probability one, as sample size increases, the sample averaged will converge to the underlying matrix of distribution means, which are probably the knowledge we seek to represent [10]. But in FCM literature only a few methods for aggregating FCM have been proposed.

The firsts and also simplest procedures were proposed by Kosko [2]. They are based on the mathematical transformation of the causal weight matrixes. If the maps are characterized by the same concepts, then the aggregated FCM that represents the entire system can be easily calculated as the average, median or weighted average of their causal matrixes.

But, usually the experts have different opinions regarding the concepts involved in the modeling. In such cases, it is necessary to augment each causal matrix by adding a new column and row filled with zeros for each additional concept. As a result, all causal matrixes have the same dimension and it is possible to perform the above mentioned operations among them. Next figure shows the aggregation process of two FCM where different concepts (problem variables) are selected for describing the investigated system.

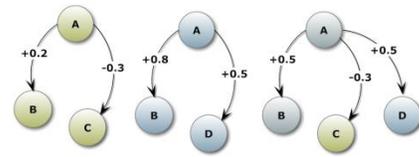


Fig.1: Typical aggregation process of FCM.

On the other hand, despite the numerous advantages of FCM knowledge representation and reasoning, tasks related to FCM modeling are challenging for humans due to the large number of aspects that should be taken into account in this process. For example, the estimation of the system causality among  $n$  concepts could involve a matrix having  $n \times n$  continuous values in range  $[0,1]$ . For addressing these drawbacks, some learning algorithms were introduced in past years, but more research is required. Thus, in next Sections we describe two abstract learning models based on Swarm Intelligence for optimizing some parameters on the FCM design: the system causality and also the map topology.

### 3. Learning system causality

In the FCM literature numerous learning algorithms to estimate the weight matrix from historical data have been proposed. Depending on the human interaction, they may be gathered in two major groups: the semi-automated and the fully-automated models. The first ones still require a relatively limited intervention, whereas fully automated approaches are able to compute a FCM model solely based on historical data [11]. Next, a fully automated model which is able to compute the system causality using historical data as learning knowledge is presented.

As a brief categorization, learning procedures for FCM could be classified into two main groups: the Hebbian approaches and the population-based algorithms, although in recent years hybrid paradigms have been successfully introduced. In this paper we adopt a population-based method, since these search methods are able to discover promising solutions in a reasonable execution time, and in general terms they are easy to implement.

From the author's point of view, designing a population-based algorithm for estimating the system causality involves three key aspects: the individual's representation, the fitness function design and also the selection of a competent optimizer according with search features.

#### 3.1. Representing individuals

As was mentioned before, the causality is gathered in a weight matrix  $W$ , where the component  $w_{ij}$  controls the interaction between the cause concepts  $c_i$  and the effect node  $c_j$ . If  $w_{ij} > 0$  then a positive causality exist, so increasing (decreasing) the value of the cause concept leads to an augment (decrement) in the activation value of the final concept. If  $w_{ij} < 0$  then an alteration on the initial concept causes an opposite effect over the target concepts, otherwise there exists no causal relation.

Regardless of the restrictions of the map topology and the selected optimizer, the causal links should be codified into a vector in order to obtain a feasible representation of individuals. For example, suppose a FCM having  $n$  concepts, where self-connections are not allowed. So, we can transform the associated weight matrix in a vector in the  $R^{n \times (n-1)}$  space where each dimension denotes a causal relation on the map. On other words, each row of the weight matrix is taken as a genotype sequence in the agent codification. However, a sensible question remains: how to efficiently design the fitness function?

#### 3.2. Designing the fitness function

The formulation of a proper fitness function is an important aspect in population-based learning methods. The equation (2) shows the general expression of proposed objective function. It is used to compute the fitness value of an individual  $x$ . Then, the quality of the candidate FCM is iteratively improved by maximizing this function during the search process, therefore minimizing the total system error, which is formalized in equation (3).

$$f(x, \varphi) = \frac{1}{1 + Er(x, \varphi)} \quad (2)$$

$$Er(x, \varphi) = \sum_{i=1}^{|\varphi|} \|\mu(x, \varphi_i), R_i\|_L \quad (3)$$

Where  $x$  is the candidate weight matrix generated by the optimizer,  $\varphi$  represents the set of historical data,  $R_i$  is the known response for the  $i$ -th input data sequence, while the function  $\mu(x, \varphi_i)$  computes the current system response, that is, the map inference for the conditions imposed for the input sequence vector. Should be also mentioned that  $\|\cdot\|_L$  refers to a norm (e.g. the  $L_1$ -norm,  $L_2$ -norm or  $L_\infty$ -norm) that is used to calculate the inference error.

#### 3.3. Selecting a competent optimizer

Selecting an accurate optimizer is another important aspect that should be severely considered in the learning algorithm design. In this paper we use a scheme based on Particle Swarm Optimization (PSO) in order to illustrate the proposed methodology. Nevertheless, other population-based search algorithms may be adopted.

The PSO meta-heuristic [12] involves a set of particles known as *swarm* which explore the search space trying to locate promising regions. A particle represents a problem solution, and it is codified as a vector in the space  $R^n$ . In the case of the canonical PSO, each particle  $x_i$  moves through the space using its own velocity  $v_i$ , a local memory of the best position it has obtained  $p_i$  and knowledge of the best solution  $g$  found in its neighborhood.

Equations (4) and (5) show how to update the velocity and position of each particle based on two factors: the *cognitive* knowledge (i.e. the local memory of best solution found by each particle) and the *social* knowledge (i.e. the global memory of the best particle found so far during the search). Here both *cognitive* and *social* components are established by using two parameters  $\delta_1 \sim U(0, c_1)$  and  $\delta_2 \sim U(0, c_2)$  where  $c_1$  and  $c_2$  should be fixed by users, although more recent variants are able to estimate them in a self-adaptive way.

$$v_i = v_i + \delta_1(p_i - x_i) + \delta_2(g - x_i) \quad (4)$$

$$x_i = x_i + v_i \quad (5)$$

Generally speaking, the features of this meta-heuristic for dealing with global optimization problems include the improved capability of solving complex problems, high convergence speed and good generality for different problems [13]. Due to the high convergence rate characterizing PSO, particle swarm is quickly attracted to stable

points, but these solutions are not necessarily global optima.

This behavior causes premature convergence of the PSO algorithm, where the swarm is crowded about local solutions with little chance of escaping from this situation [14, 15]. In fact, once particles fall in a premature convergence situation, they continue converging within extremely close proximity of one another so that the global best and all personal bests are within one small region of the space, seriously limiting the search exploration.

In the context of FCM optimization, it is notable that the heuristic functions used for tuning FCM parameters regularly are not convex and hence they are characterized by multimodal search spaces where multiple optima solutions exist; causing premature convergence of learning algorithms to sub-optimal solutions. In summary, a good optimizer should be capable to handle the following problems:

- **High dimensionality:** in many application problems the number of concepts/relations is quite large, thus the optimizer should enclose good scalability features to deal with.
- **Premature convergence and stagnation states:** typically objective functions are multimodal in nature having many local as well as global optima, and hence the optimization algorithm may fall in local optima, seriously affecting the learning performance.

In this work we adopt an improved variant of PSO using a simple but powerful method called Random Sampling in Variable Neighborhoods (PSO-RSVN), which is capable to outperform other PSO algorithms [15]. The main idea of the PSO-RSVN is to restructure the particle swarm from the selection of a set of samples in several neighborhoods, once stagnation or convergence premature states are detected. Besides, it improves the exploration as well as exploitation abilities of the PSO during the search process.

#### 4. Learning central concepts

Most learning algorithms reported in the FCM literature are oriented to the estimation of the system causality. Less attention have been given to the map concepts, however, it could occur that the variables have not the same importance or are superfluous, thus, their incorporation in the model instead of benefiting may affect the model interpretability.

This situation also seriously affects the FCM aggregation procedure described in the Section 2.1. When multiple maps are aggregated, usually all different variables from multiple experts are present in the augmented map; while the causal values concerning to the edges are combined through an operator. However, it is possible that a subset of the systems variables is not important in the final modeling, or even they could represent contradictory

concepts, which strongly affect the aggregated FCM performance.

For example, suppose that  $m$  maps will be aggregated into a single one, and each map have a different superfluous concept (i.e. its influence on the system inference is very weak). Thus the augmented map could have  $m$  superfluous variables, and thus could have at most  $m^2$  non-relevant causal links. Clearly, these concepts and relationships difficult the system interpretation and may affect the map inference.

Finding central concepts in a FCM involves a complex problem, and its solution using direct methods could be a challenge, mainly due to the large number of time-consuming operations. For addressing this drawback, next we describe an optimization model to determine central nodes in the net, by using the Ant Colony Optimization (ACO) meta-heuristic. Similarly to section before, we need to specify the individual's representation, the fitness function design and also to justify the proper selection of the optimizer used in the search process.

##### 4.1. Representing individuals

The first step for finding central concepts in a FCM using a population-based method is to represent solutions. In this way we use a binary vector with cardinality  $n$  (denoting the number of variables in the system). Thus, a zero in the  $i$ -th component of a candidate solution vector means that the concept  $C_i$  will be removed from the initial modeling; otherwise the node will be preserved in the optimized model.

Then the goal of the optimizer is to find the binary vector having the lowest norm, which preserve the quality in the inference process. But, it is important to remark that in some kind of problems there exist variables directly related with the system modeling, and they cannot be removed from the map. These restrictions should be considered in the optimization process, and of course they depend on the application domain.

##### 4.2. Designing the fitness function

Once the specifications of the solutions have been properly formalized, it is essential to define the fitness function. Here we use a function  $0 \leq g < 1$  composed by two factors. The first factor quantify the relation between the number of concepts removed respect to the total number of system variables; while the second factor is oriented to maximize the quality of the map simulations, by reducing the overall error as the formula (2) shows.

$$g(y) = \frac{(1 - \lambda)f(x, \varphi) + \lambda |\phi_y|}{2} \quad (6)$$

In addition,  $0 \leq \phi_y \leq 1$  represents the number of non-relevant concepts, respect to the total number of variables involved, while the function  $f$  measures the system re-

sponse quality of the map for the current causal weight matrix  $x$ , and the dataset of input sequences stored in the knowledge base  $\varphi$  (see Section 3.2).

In this expression we introduce a new parameter  $0 < \lambda < 1$ . This factor regulates the importance that the user confers to the quality in the FCM inference, respect to the number of removed variables. In this context, the authors suggest that  $\lambda < 0.4$  since the main idea of the learning algorithm is to find central concepts by reducing the number of non-relevant variables, but always preserving the system modeling features and hence the strength in the FCM inference capabilities.

### 4.3. Selecting a competent optimizer

The main goal of the discrete optimizer is to generate a good enough binary solution ensuring a reduction on the map cardinality, and preserving the quality in the inference process. To achieve this, we use the ACO meta-heuristic as a scheme for explain our proposal. But, why ACO can be considered a good optimizer for finding central concepts in a FCM?

The ACO meta-heuristic is a stochastic search method [16] which is based on the shortest path searching behavior observed in ant specie. When ants search for food in nature, they originally explore the area following a random pattern. As soon as an ant finds a food source it evaluates the quality as well as the quantity of the food. Hence, during the return tour, the ant deposits a chemical pheromone trail on the ground. This trail will guide other ants to the discovered food source.

The ACO algorithm is a fully constructive model which is inspired in this behavior, where each ant builds a solution of the combinatorial problem by traveling a construction graph. The next graph node that should be visited (i.e. the next solution component) is stochastically selected by using two kind of knowledge: the pheromone trail, and the heuristic information. Following equation (7) shows the relative probability of the  $k$ -th ant of moving from node  $i$  to node  $j$  when the ant travels through the edge  $a_{ij}$ .

At this point, it is important to mention that the heuristic knowledge provides an estimation of ant's preference of moving from current node  $i$  to node  $j$ . This information is known in advanced and it is not modified by ants during the search process. On the other hand, the artificial information measures the learned desirability of agent's movements and it is iteratively updated mainly depending on the fitness of discovered solutions.

$$P_{ij}^k(t+1) = \frac{[\tau_{ij}(t)]^\alpha [\eta_{ij}]^\beta}{\sum_{r \in N_i^k} [\tau_{ir}(t)]^\alpha [\eta_{ir}]^\beta} \quad (7)$$

In the above formula  $t$  denotes the current step,  $\tau_{ij}$  is the pheromone trail information, and  $\eta_{ij}$  is the heuristic information, while the set  $N_i^k$  represents the feasible neighborhood of the  $k$ -th ant. Also, the ACO model involves

two parameters  $\alpha$  and  $\beta$  that should be specified by users: the relative influence of the pheromone trail and the heuristic information.

If  $\alpha = 0$ , the selection probabilities are proportional to the factor depending on the heuristic information, and thus the closest states will more likely be selected by the ants. On other hand, if  $\beta = 0$ , only pheromone amplification is working; this will lead to the quick emergence of a stagnation situation [17], frequently falling into strong local solutions.

After the solution construction is done, ants update the pheromone trail. First, all pheromone values are uniformly reduced in the evaporation step. Afterward, one or more solutions are used for increasing the values of the pheromone trail involved on these solutions. Actually, different ACO variants/algorithms mainly differ in the pheromone strategy that they apply.

We adopted the Max-Min Ant System (MMAS) since it provides a stronger exploration, besides it attempts to avoid the search stagnation more effectively [18]. Following equation (8) shows the pheromone updating strategy used in MMAS.

Then, after each algorithm cycle, the evaporation process will reduce the trail strength using a constant factor  $0 < \rho < 1$ , and next only the best ant is allowed to modify the pheromone trail in each cycle. However, the strong elitism features induced by this rule may leads to the model being trapped in a local solution.

$$\tau_{ij}(t+1) = (1-\rho)\tau_{ij}(t) + \rho\tau_{ij}^b \quad (8)$$

Reducing the undesirable effects of the search stagnation, explicit limits for all pheromone trails are established by using two parameters  $\tau_{min}$  and  $\tau_{max}$ . In each step must be ensured that pheromone trails respect these limits. Besides, all pheromone trails are initialized as  $\tau_{max}$ , achieving better exploration of the search space at the algorithm beginning [19].

From the point of view of the selection of central concepts in a FCM, it is notable the "role" that should play a proper estimation of the heuristic factor. Following equation (9) introduces a heuristic function based on a centrality measure. To do so, we use the centrality degree because it is a very simple local measure for computing the graph centrality, which it is determined by only directed links.

$$\eta_i = \sum_{i=1}^m \frac{|w_{il}|}{m-1} + \sum_{p=1}^m \frac{|w_{pi}|}{m-1}, \forall i \quad (9)$$

In the above formula,  $m$  is the concepts number involved in the investigated system,  $w_{il}$  and  $w_{pi}$  denote causal weights associated with the concept  $c_i$ , whereas  $m-1$  represents the maximal number of causal relations that could be linked with each concept, assuming that self-connections are not allowed, and it is used to normalize the heuristic value. However, in some problems it is not

possible to obtain fully-connected maps, thus this value should be adjusted for each problem, according to the FCMs topology restrictions.

In the next sections we describe two study cases for illustrating the utility of the learning algorithms proposed. The first study case is related to a bioinformatics problem where the learned causality represents the influence of a protein position in the phenotypic resistance. The second study case is related to travel behavior in developed cities, where an optimization of the number of the system variables is necessary to facilitate the system interpretation and usability.

### 5. First case study: discovering biological causality in a HIV protein

It is well-known that the Human Immunodeficiency Virus causes millions of deaths yearly around the world. Although several drugs are approved for treating this disease, they are unable to eradicate the virus since these antiviral are mainly designed for reducing the virus concentrations. Specifically, these drugs try to inhibit the function of three essential proteins which are relevant in the virus life cycle: *protease*, *integrase*, *reverse transcriptase*. For example, the last enzyme is related with the reverse transcription process replicating the genetic information in the host cell and consequently releasing new infected viral particles.

Usually, these antivirals are successfully combined in a highly active antiretroviral therapy. However, due to its complexity and high mutation rate, the HIV is capable to develop resistance, eventually causing the therapy failure. For this reason, it is necessary a deeper study of resistance mechanisms with the intention of make a better use of existing drugs.

The resistance testing of a mutation can be performed by two main methods. The phenotypic test which measures the viral activity in presence of a drug concentration, and the genotypic test which scans the mutation sequence for certain mutations previously associated with resistance [20]. The results of both tests offer a relevant source of information about the relation between the genetic mutations and the resistance to a given drug. Consequently, several statistical and machine learning techniques have been proposed for predicting the phenotype from the genotypic information [20-22].

In recent works, the modeling of this bioinformatics problem using FCM theory allowed discovering causality patterns on the *protease* protein, as well as central amino acids [5, 23]. Next we will use this model and the continuous learning algorithm for discovering causality between

positions of the *reverse transcriptase* protein and the drug resistance node.

#### 5.1. Protein modeling using FCM

Here, the *reverse transcriptase* protein is described by a sequence of 560 positions, where each one of them is one of the 20 possible amino acids. However, since not all positions are directly related to the resistance, we use a feature selection of protein positions previously associated with resistance from the biological point of view for representing this protein [21]. In addition, this feature selection improves the interpretability of the final model by reducing its dimensionality.

Another aspect taken into account is the features description. In this work, all protein positions are characterized by its contact energy [24], which is a numerical descriptor corresponding, to some extent, to the tridimensional structure of the protein and it is calculated statistically from a large number of diverse sequences.

Figure 2, illustrates how each protein position is expressed as a system variable (concept) and another extra concept for representing the drug resistance is added to the model. Also, causal relations are established among protein positions and between each position and the resistance expressing the causal influence and interactions taking place in the protein.

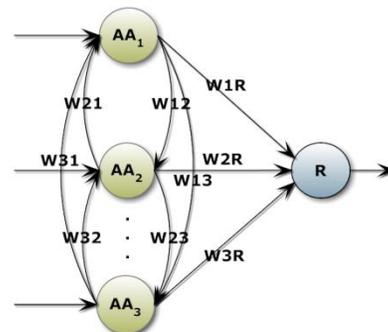


Fig.2: Topology for HIV protein modeling.

Then applying the continuous learning algorithm previously described, it is possible to discover the underlying biological patterns characterizing the behavior of the HIV *reverse transcriptase* protein.

#### 5.2. Knowledge discovering

Using historical data from the Stanford HIV Resistance Database [23] six new knowledge bases are built for representing the reported resistance of mutations to the following nucleoside *reverse transcriptase* inhibitors: Lamivudine (3TC), Abacavir (ABC), Stavudine (D4T),

Table 1: Prediction accuracies obtained after the learning process over 10 runs.

Drug	LSR	SVR	LARS	DT	NN	FCM
3TC	0.83	0.84	0.88	0.90	0.90	0.79
ABC	0.63	0.65	0.77	0.69	0.66	0.71
AZT	0.64	0.70	0.76	0.70	0.71	0.71
D4T	0.66	0.68	0.78	0.75	0.72	0.90
DDI	0.61	0.67	0.75	0.74	0.71	0.83
TDF	0.46	0.69	0.70	0.68	0.73	0.84

Table 2: Amino acid causal influences over resistance for D4T

AA41	AA44	AA62	AA65	AA67	AA69	AA70	AA74	AA75
-0.1137	-0.1015	-0.1455	-0.1397	-0.1591	0.0217	-0.1654	0.1964	0.0225
AA77	AA115	AA116	AA118	AA151	AA184	AA210	AA215	AA219
-0.0455	0.0200	0.1968	-0.1550	0.0767	0.1114	0.0529	-0.2285	-0.0891

Didanosine (DDI), Tenofovir (TDF) and Zidovudine (AZT). From 10 trials we obtained equal number of adjusted maps for each drug, using as parameter settings for the learning algorithm: 80 particles, 200 generations, five variable neighborhoods, and a tolerance (threshold) for the maximum radius of the swarm  $\alpha = 1.0E-5$ .

The prediction accuracies obtained in a 5-folds cross-validation of the learning process are presented in the Table 1. The models used for comparison include: linear regression (LSR), support vector regression (SVR), linear discriminant analysis (LARS), decision trees (DT) and neural networks (NN) taken from [22] using the complete set. The results obtained are competitive with other techniques and they also are, to some extent, a quality measure of the knowledge expressed in the causal relations.

After the learning process is done, the following step is to perform an aggregation of several adjusted maps obtained for each drug (described in Section 2.1). The aggregation procedure is a proper tool for making a consensus among the different weight matrixes found by the optimization algorithm. As a result we obtained a prototype map for each drug representing the final biological causality.

Just then, using the prototype map as a knowledge source we can discover causal patterns among protein positions and the resistance. Next table 2 shows the discovered patterns on the prototype map corresponding to Stavudine (D4T).

For the positions 74, 116, 151, 184 and 210 a positive causality exists, it means that a mutation to an amino acid having more contact energy on these positions leads to an increase in the resistance to the drug. Also for the positions 41, 44, 62, 65, 67, 70, 77, 118, 215 and 219, a negative causality exists; so a change in the contact energy value inversely affects the resistance. Finally, no significant causality is observed for amino acids 69, 75 and 115. However, it is important to remark that a single mutation on these positions could be significant for the enzyme, considering all amino acids interactions.

With the same criteria, it can be drawn conclusions about causal patterns in the direct relations of the other

five modeled drugs. This discovered knowledge characterizes, in part, the interactions that occur in the active site of the protein between the enzyme and its substrate, allowing designing inhibitors that avoid the enzymatic reaction, leading to produce immature and non-infectious virions.

Although we only investigate the direct causality between a protein positions and the resistance, the modeling of the protein behavior involves in its representation scheme the information for simulating the global effects of multiple mutations over the wild genome. Varying the contact energies of inputs in a trained map and activating the inference process is enough for interpreting the causal interactions.

For example, a drug expert could predict the resistance to an antiviral when multiple mutations take place and influence the resistance; besides, it is possible to visualize how is affected the resistance concept by positions that have been stimulated by other mutations. These simulations could help drug designers and other specialists to make a better use of existing drugs and also study the resistance mechanisms of this complex protein.

## 6. Second case study: travel behavior analysis using FCM theory

The second application domain is related to the analysis of public transportation on modern societies. Here, the selection of a specific transport mode by a user is conditioned by an extensive range of context situations. Thus, analyzing the travel behavior allows to policy-makers the formulation of new strategies that efficiently manage existing resources and also correspond with customer's preferences.

Transport demand problems are non-structured by nature and normally they include imprecise descriptions, vagueness or uncertainty. Therefore the FCM theory seems to be an appropriate choice. Recently, M. León et al. [25, 26] proposed a novel modeling based on Cognitive Mapping theory to characterize mental representations of transportation experts.

As well, a learning algorithm based on Swarm Intelligence for tuning the initial causality among concepts is discussed, allowing simulating underlying patterns.

### 6.1. Proposed FCM topology

Clearly, transport behavior problems are characterized by a wide set of context variables (e.g. temperature, physical condition, parking cost, car availability, fuel cost, traffic control, bus frequency, reliability, travel time, bus cost, etc.) depending of the user requirements. The proposed model includes three kinds of variables which are straightly translated as FCM nodes: *situation* variables, *attribute* variables, and *benefit* variables.

Here *situation* variables denote input concepts describing environmental situations, whereas *attribute* and *benefits* ones are receiving concepts associated to situation variables, that is, they are only influenced by other concepts. Figure 3 shows the proposed topology, describing the interaction among system variables.

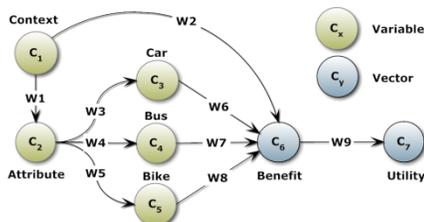


Fig.3: FCM for travel behavior analysis.

Besides the above concepts, three further concepts for each transport mode are considered. Hence when the FCM inference process (propagation of the concept activation level) is activated, a map node called “Utility” is examined. It has evidence related with the user preferences respect to the available transport modes (*car*, *bus* and *bike*) for an initial condition fixed for input or situation concepts.

Notice that *benefits* concepts as well as the “Utility” node are encoded as vectors. From the machine learning point of view, the proposed topology leads to a complex multi-objective regression problem, where the expected utility that reports the selection of a specific transport mode should be estimated. However, as a result from [26], we know that some concepts are severely susceptible to expert’s criteria and they are often superfluous or contradictory. Then, finding central nodes could reduce the map complexity, allowing the knowledge extraction tasks.

### 6.2. Finding central concepts

In order to explain the process of finding central concepts we use 10 maps obtained from [27], where each map represents an expert. It is remarkable that, the average number of *attribute* nodes is around 34, while the average number of causal links is approximately 122. Therefore, the proposed learning algorithm has to explore a large

search space having  $2^{34}-1$  feasible solutions, which confirms the high complexity of this combinatorial problem.

The values for the parameters used for MMAS are: 10 ants,  $\rho = 0.1$ , the constant of evaporation,  $\alpha = 2$  and  $\beta = 3$ , whereas the pheromone thresholds  $\tau_{min}$  and  $\tau_{max}$  are dynamically calculated as suggest reference [18]. Finally, the relative importance conferred to the inference quality over the number of removed nodes is  $\lambda = 0.95$ .

Table 3 shows a summary of the simulation results over 10 trials. The semantic of these experiments are explained as follows. As a first step, the number of nodes and causal connections are calculated. It allows quantifying the overall density of selected maps without any modification. Consequently, the discrete learning algorithm is performed and the density of the global best map is analyzed.

In general terms the reader should notice that the number of concepts decreases on 18.47%, while the reduction on the causal links achieves 26.41%. From these results an important conclusion came out: the discrete learning algorithm discussed in this paper not just is able to reduce the number of variables or concepts involved in the investigated system, but also it is able to reduce the global map density. As a result, optimized maps are simpler and consistent from the structural perspective, thus allowing better interpretation.

Table.3: Influence of discrete learning method over the relative density of selected maps.

map	Before		After	
	nodes	links	nodes	links
1	37	145	28	103
2	35	126	28	88
3	37	112	31	86
4	41	155	34	109
5	32	123	26	88
6	35	126	28	94
7	37	138	34	126
8	27	100	21	67
9	30	100	25	71
10	30	98	23	68

## 7. Conclusions

The FCM theory is a suitable Soft Computing technique for modeling and simulation of dynamical and complex systems, but establishing their parameters could be a challenging task for humans. In this paper was presented two learning schemes based on Swarm Intelligence for correcting the parameters of these structures.

The first learning algorithm adjusts the causal weight matrix only using historical data, while the second one finds the central concepts in the map removing superfluous or contradictory features. In both proposals the authors designed the individual’s representation, as well as the fitness function. Conversely, we suggested and justi-

fied the proper use of adopted optimization algorithms for this purpose: PSO-RSVN and MMAS.

In order to perform the proposals, we introduce two study cases. The first one is related to the analysis of the HIV *reverse transcriptase* enzyme, where causal patterns among all protein positions and the resistance are learned through the weights adaptation methodology. As a result, the prediction accuracies obtained were competitive with respect to other techniques taken from literature. Besides, the interpretability of this FCM allowed discovering knowledge that could be useful for designing more effective therapies, and simulating the effects of mutations.

The second real study case is related to travel behavior problems in modern societies, where the user's choice respect to a specific transport mode regularly depends of numerous variables. But, some of them are often redundant or contradictory. To find central concepts in a map, we introduced a learning method which is able to reduce the net complexity without lost the system knowledge. As a result, optimized maps allow to policy-makers better understanding of the system for efficiently managing the existing resources, and also correspond with customer's preferences.

Of course the proposed learning methodology for tuning both discrete and continuous parameters in a FCM is still far to be perfect. Actually, more research work needs to be done toward the extension of the learning capabilities for enhancement and adaptation of the FCM structures, as well as for establishment of new learning formalisms to overcome explicit weaknesses [28]. Future work will be focused on presenting a new hybrid learning algorithm combining a more scalable Cooperatively Particle Swarm with Hebbian-type approach, in order to improve the overall precision of FCM parameters.

## 8. References

- [1] R. Axelrod, "Structure of Decision: The Cognitive Maps of political Elites," *Princeton University Press*, 1976.
- [2] B. Kosko, "Fuzzy Cognitive Maps," *International Journal of Man-Machine Studies*, vol. 24, pp. 65-75, 1986.
- [3] B. Kosko, "Neural Networks and Fuzzy systems, a dynamic system approach to machine intelligence," *Prentice-Hall*, 1992.
- [4] M. León, G. Nápoles, et al., "Two steps Individuals Travel Behavior Modeling through Fuzzy Cognitive Maps Pre-definition and Learning," *Lecture Notes in Computer Sciences*, vol. 7095, pp. 82-94, 2011.
- [5] G. Nápoles, I. Grau, M. León and R. Grau, "Modelling, aggregation and simulation of a dynamic biological system through Fuzzy Cognitive Maps," *Lecture Notes on Artificial Intelligence*, vol. 7630, pp. 188-199, 2013.
- [6] E.I. Papageorgiou, "Review study on Fuzzy Cognitive Maps and their applications during the last decade," *Proc. IEEE International Conference on Fuzzy Systems*, pp. 828-835, 2011.
- [7] D.E. Koulouriotis, D.M. Diakoulakis, D.M. Emiris, E.N. Antonidakis and I.A. Kaliakatsos, "Efficiently modelling and controlling complex dynamics systems using evolutionary Fuzzy Cognitive Maps (Invited paper)," *Int. J. Comput. Cognition*, vol. 1, pp. 41-65, 2003.
- [8] B. Kosko, "Fuzzy Thinking," 1993/1995, (Chapter 12: Adaptive Fuzzy Systems).
- [9] A.K. Tsadiras, "Comparing the inference capabilities of binary, trivalent and sigmoid fuzzy cognitive maps," *Information Science*, vol. 178, pp. 3880-3894, 2008.
- [10] B. Kosko, "Hidden patterns in combined and adaptive knowledge networks," *International Journal of Approximate Reasoning*, vol. 2, pp. 377-393.
- [11] M. León, G. Nápoles, et al., "A Fuzzy Cognitive Maps Modeling, Learning and Simulation Framework for Studying Complex System," *Lecture Notes on Artificial Intelligence*, vol. 6687, pp. 243-256, 2011.
- [12] J. Kennedy and R. Eberhart, "Particle Swarm Optimization," *Proc. of the 1995 IEEE International Conference on Neural Networks*, pp. 1942-1948, 1995.
- [13] Y. Wang, B. Li, T. Weise, J. Wang, B. Yuan and Q. Tian, "Self-adaptive learning based particle swarm optimization," *Information Sciences*, vol. 181, pp. 4515-4538, 2011.
- [14] G. Nápoles, I. Grau and R. Bello, "Particle Swarm Optimization with Random Sampling in Variable Neighbourhoods for solving Global Minimization Problems," *Lecture Notes in Computer Sciences*, vol. 7461, pp. 352-353, 2012.
- [15] G. Nápoles, I. Grau and R. Bello, "Constricted Particle Swarm Optimization based algorithm for global optimization," *POLIBITS Research journal on Computer science and computer engineering with applications*, vol. 46, pp. 5-11, 2012.
- [16] M. Dorigo, V. Maniezzo and A. Colomi, "The Ant System: Optimization by a colony of cooperating agents," *IEEE Transactions on Systems, Man, and Cybernetics*, Part B, vol. 26, pp. 1-13, 1996.
- [17] M. Dorigo, E. Bonabeu and G. Theraulaz, "Ant algorithms and stigmergy," *Future Generation Computer Systems*, vol. 16, pp. 851-871, 2000.
- [18] T. Stutzle and H. Hoos, "MAX-MIN ant system," *Future Generation Computer Systems*, vol. 16, pp. 889-914, 2000.
- [19] A. Puris, R. Bello and F. Herrera, "Analysis of the efficacy of a Two-Stage methodology for ant colony optimization: Case of a study with TSP and QAP," *Expert system with Applications*, vol. 37, 543-553, 2010.
- [20] N. Beerenwinkel, B. Schmidt, H. Walter, R. Kaiser, T. Lengauer, D. Homann, K. Korn, and J. Selbig, "Diversity and complexity of HIV-1 drug resistance: a bioinformatics approach to predicting phenotype from

- genotype,” *Proceedings of the National Academy of Sciences of the United States of America*, vol. 99, pp. 8271-8276, 2002.
- [21] S.-Y. Rhee, M.J. Gonzales, R. Kantor, B.J. Betts, J. Ravela and R.W. Shafer, “Human immunodeficiency virus reverse transcriptase and protease sequence database,” *Nucleic Acids Research*, vol. 31, pp. 298-303, 2003.
- [22] S.-Y. Rhee, J. Taylor, G. Wadhera, A. Ben-hur, D.L. Brutlag and R.W. Shafer, “Genotypic predictors of human immunodeficiency virus type 1 drug resistance,” *Proceedings of the National Academy of Sciences of the United States of America*, vol. 103, pp. 17355-17360, 2006.
- [23] G. Nápoles, I. Grau, R. Bello and R. Grau “Two-steps learning of Fuzzy Cognitive Maps for prediction and knowledge discovering on the HIV-1 drug resistance,” *Expert Systems with Applications*, 2013 (submitted).
- [24] S. Miyazawa and R.L. Jernigan, “Contacts energies Self-Consistent Estimation of Inter-Residue Protein Contact Energies Based on an Equilibrium Mixture Approximation of Residues,” *PROTEINS: Structure, Function, and Genetics*, vol. 34, pp. 49-68, 1999.
- [25] M. León, G. Nápoles, et al., “A Revision and Experience using Cognitive Mapping and Knowledge Engineering in Travel Behavior Sciences,” *POLIBITS Research journal on Computer science and computer engineering with applications*, vol. 42, pp. 43-49, 2010.
- [26] M. León, G. Nápoles, R. Bello, L. Mkrtychyan, B. Depaire, K. Vanhoof, “Tackling Travel Behaviour: An approach based on Fuzzy Cognitive Maps,” *International Journal of Computational Intelligence Systems*. (in press), 2013
- [27] M. León, “A Fuzzy Cognitive Maps Based Model for Individual Travel Behaviour,” Ph.D. Thesis, University of Hasselt, 2012.
- [28] E.I. Papageorgiou, “Learning Algorithms for Fuzzy Cognitive Maps - A Review Study,” *IEEE Transactions on Systems, Man, and Cybernetics*, vol. 42, pp. 828-835, 2012.