

Learning stability features on sigmoid Fuzzy Cognitive Maps through a
Swarm Intelligence approach

Peer-reviewed author version

NAPOLES RUIZ, Gonzalo; Bello, Rafael & VANHOOF, Koen (2013) Learning stability features on sigmoid Fuzzy Cognitive Maps through a Swarm Intelligence approach. In: Ruiz-Shulcloper, José; Sanniti di Baja, Gabriella (Ed.). Progress in Pattern Recognition, Image Analysis and Applications: 18th Iberoamerican Congress, CIARP 2013, Havana, Cuba, November 20-23, 2013, Proceedings, Part I, p. 270-277.

Handle: <http://hdl.handle.net/1942/21460>

See discussions, stats, and author profiles for this publication at: <https://www.researchgate.net/publication/258332737>

Learning Stability Features on Sigmoid Fuzzy Cognitive Maps through a Swarm Intelligence Approach

Conference Paper · November 2013

DOI: 10.1007/978-3-642-41822-8_34

CITATIONS

4

READS

97

3 authors:



[Gonzalo Nápoles](#)

Hasselt University

31 PUBLICATIONS 124 CITATIONS

SEE PROFILE



[Rafael Bello](#)

Central University "Marta Abreu" of Las Villas

161 PUBLICATIONS 628 CITATIONS

SEE PROFILE



[Koen Vanhoof](#)

Hasselt University

243 PUBLICATIONS 1,998 CITATIONS

SEE PROFILE

Learning stability features on sigmoid Fuzzy Cognitive Maps through a Swarm Intelligence approach

Gonzalo Nápoles^{1,*}, Rafael Bello¹ and Koen Vanhoof²

¹Universidad Central “Marta Abreu” de Las Villas, Santa Clara, Cuba

²Hasselt University, Diepenbeek, Belgium

*gnapoles@uclv.edu.cu

Abstract. Fuzzy Cognitive Maps (FCM) are a proper knowledge-based tool for modeling and simulation. They are denoted as directed weighted graphs with feedback allowing causal reasoning. According to the transformation function used for updating the activation value of concepts, FCM can be grouped in two large clusters: discrete and continuous. It is notable that FCM having discrete outputs never exhibit chaotic states, but this premise can not be ensured for FCM having continuous output. This paper proposes a learning methodology based on Swarm Intelligence for estimating the most adequate transformation function for each map neuron (concept). As a result, we can obtain FCM showing better stability properties, allowing better consistency in the hidden patterns codified by the map. The performance of the proposed methodology is studied by using six challenging FCM concerning the field of the HIV protein modeling.

Keywords: Fuzzy Cognitive Maps, Stability, Learning, Swarm Intelligence.

1 Introduction

The FCM theory [1] was introduced by B. Kosko as an improvement of the Cognitive Mapping which uses fuzzy reasoning in its knowledge representation scheme. From the structural point of view, FCM may be denoted as directed graphs with feedback, consisting of nodes and weighted arcs. Nodes or concepts are equivalent to neurons in connectionist models and represent variables of the modeled system; while weights associated to links denote the *causality* among concepts. In a FCM, each connection takes value in the range $[-1,1]$. It denotes the degree of causality between two nodes as a result of the quantification of a fuzzy linguistic variable [2-3], which is regularly assigned by experts at the modeling stage. The activation value of concepts is also fuzzy in nature and regularly takes values in the range $[0,1]$, although the scale $[-1,1]$ is also allowed. Hence, the higher the activation value, the stronger the influence of the concept over the system, which helps to understand the modeling.

In the past decade, FCM have gained considerable research interest and are widely used to analyze causal systems such as system control, decision making, management, risk analysis, text categorization, prediction, etc [4]. In order to increase the usability of this knowledge-based approach, numerous researchers have developed learning algorithms [3], mainly varying the causal weight matrix. However, these approaches suppose that FCM are closed systems and they do not consider external influences,

while other factors such as the FCM stability are ignored. On the other hand, as far as known, there not exist learning algorithms for enhancing the system stability once the causality is established. Based on these considerations, the main goal of this work is to introduce a new learning algorithm which is oriented to estimate the most adequate transformation function for each map concept, simulating the effect of ideal external stimulus over the neurons with the hope to improve the map stability. To do that, we use a Swarm Intelligence method to solve the related optimization task.

It is relevant to remark that the authors will be focused on sigmoid FCM, instead of discrete (binary or trivalent) maps. This remark is motivated by the benchmarking analysis discussed in [5] where results showed that the sigmoid function significantly outperforms the other functions, by using the same decision model. The rest of the paper is organized as follows: in next Section 2 the formulation of FCM is briefly described. In Section 3 we present the proposed learning algorithm which is oriented to compute the family of threshold functions improving the map convergence. After that, Section 4 introduces the experimental framework and also provides comments about the simulations. Finally, conclusions are given in Section 5.

2 Fuzzy Cognitive Maps

Without loss generality, a simple FCM can be defined using a 4-tuple (C, W, A, f) where $C = \{C_1, C_2, C_3, \dots, C_m\}$ is a set of m concepts of the graph, $W: (C_i, C_l) \rightarrow w_{il}$ is a function which associates a causal value $w_{il} \in [-1, 1]$ to each pair of nodes (C_i, C_l) , denoting the weight of the directed edge from C_i to C_l . In this context, it is important to notice that w_{il} represents the causality degree between the concepts C_i and C_l . Thus the weigh matrix $W_{m \times m}$ gathers the system causality which is frequently determined by experts. Likewise, $A: (C_i) \rightarrow A_i$ is a function that associates the activation degree $A_i \in \mathbb{R}$ to each concept C_i at the moment t ($t = 1, 2, \dots, T$). Finally, a transformation or threshold function $f: \mathbb{R} \rightarrow [0, 1]$ is used to keep the activation value of concepts in the interval $[0, 1]$. Following equation (1) shows the inference mechanism using the state vector A^0 as the initial configuration. This inference stage is iteratively repeated until a *hidden pattern* [6] or a maximal number of iterations T are reached.

$$A_{it+1} = f\left(\sum_{j=1}^m w_{ji} A_{jt}\right), i \neq j \quad (1)$$

The most frequently used threshold functions [5] are: the bivalent function, the trivalent function, and the sigmoid variants. The effects on the selection of a specific function over the stability and inference capabilities of the FCM have been widely explored in [7]. From this work some important remarks were concluded:

- Discrete FCM (using a binary or trivalent function) never show chaotic behavior. It means that always a fixed-point attractor or a limit cycle will be detected in the map outputs. These states have the following behavior:
 - Fixed-point attractor ($\exists t_k \in \mathbb{N} : A^{(t+1)} = A^{(t)}, \forall t > t_k$): the system produces the same output after the time t_k .

- Limit cycle ($\exists t_k, P \in \mathbb{N} : A^{(t+P)} = A^{(t)}, \forall t > t_k$): the same output or state vector is regularly observed with period P .
- Continuous FCM (using a sigmoid function) additionally may exhibit chaotic states, where the FCM model continues to produce different state vectors for successive cycles. In these situations the FCM can not stabilize.

In summary, from the algebraic point of view, the states characterizing the system are iteratively updated by multiplying the causal weight matrix, by the current state vector until a stopping condition is reached. Then the activation value of each concept is directly influenced by the values of the connected concepts with the appropriate weights, and also taking into account its previous value; showing the causal effect of changes on the concept's activation value on the whole map. That's why FCM theory is a suitable approach for handling modeling and simulation tasks.

3 Proposed learning methodology

In the literature several supervised and unsupervised learning algorithms have been proposed, mainly focused on the transformation of the causal weight matrix. As a brief categorization they can be gathered in three major groups [4]: Hebbian-based, population-based and hybrid approaches. On the other hand, Tsadiras[7] demonstrated that the inference capability of FCM may be strongly influenced by the selection of the concept's transformation function. Based on the Tsadiras' work, we conducted a set of empirical experiments using sigmoid FCM, where the parameter λ was changed as next equation (2) shows. Observe that in our simulations a custom amplification value for each map concept C_i is assumed. Results were quite promising: we observed that variations of factor λ_i lead to some changes on the map stability. Hence, it seems to be reasonable to suppose that a learning algorithm could help to improve the map convergence, by solving the related real-parameter optimization problem.

$$f_i(x) = \frac{1}{1 + e^{-(x-0.5)\lambda_i}} \quad (2)$$

Before presenting the learning methodology, we need to answer the following question: how is affected the FCM inference mechanism by the inception of a function f_i for each node C_i ? Normally FCM are considered as closed systems where external factors affecting the concepts are omitted. But, in many real world problems this perception will be inadequate and may affect the accuracy of simulations. For example, it is known that biological behavior on proteins not only depends on the amino acids interaction, but also depends on the external factors such as the chemical processes influencing the catalytic responses. Notice that such external factors may be modeled by using a function f_i for each neuron C_i (instead of the same function for all the neurons), ensuring better stability. It means that the activation value of a neuron C_i on the map will be now conditioned by the free interaction of the connected nodes and also by the steepness λ_i of its threshold function f_i which denotes the stimulus.

It should be mentioned that other works have been proposed to simulate external influence over the map concepts. For example, Stylios and Groumpos [8] introduced a new FCM model where each concept has an external output (bias), which influences each node with a weight and it is taken into account at the calculation rule. It is easy to perceive the similarity between the Stylios' model and our proposal, since both are oriented to simulate external influences over the neurons, although they use different implementations to do that. Of course, our approach leads to a different interpretation of the causal influences since the activation capability of each neuron (now influenced by its steepness λ_i) should be also considered. However, this aspect is not discussed in the paper since the authors prefer to be focused on the system stability.

Here, the learning step is focused on estimating an appropriate family of sigmoid functions ensuring certain stability features. More explicitly, we need to find a family of sigmoid functions $\{f_1(x), f_2(x), \dots, f_m(x)\}$, where the i th function will be used for transforming the activation value of the i th concepts. In practice, it implies to find the steepness λ_i for each threshold function. With this purpose in mind, we use a Swarm Intelligence technique [9]. Particle Swarm Optimization (PSO) is a non-direct search method for solving challenging continuous problems [10], from a distributed point of view without any centralized control. In the standard PSO each agent (called particle) denotes a m -dimensional point in the solution space. For our optimization problem, m should be considered as the total number of nodes on the map. Hence, the i th position of each particle will correspond to the steepness λ_i of the i th function.

Particles adjust their position by using a combination of an attraction to the best solution that they individually have found, and an attraction to the best solution that any particle has found [3], imitating those who have a better performance. This search method has proven to be quite efficient for solving real-parameter optimization problems. However, the particle swarm is frequently attracted to local optima, causing premature convergence or stagnation configurations. For this reason, this paper uses a variant called PSO with Random Sampling in Variable Neighborhoods [11-12] which is capable to notably outperform the standard algorithm. Next equation (3) shows the objective function that should be minimized during the search steps.

$$\phi(f_1, f_2, f_3, \dots, f_m) = \sum_{k=1}^K \sum_{i=1}^m \sum_{t=2}^T |A_{it}^k - A_{it-1}^k| \quad (3)$$

In the above function, K denotes the number of instances (historical data), m is the number of neurons, T denotes the maximal number of times, whereas A_{it}^k represents the activation value of the i th concept for the current time t , using the k th instance as initial condition. Here a simple instance is a sequence of values codifying the initial conditions of the system, and the corresponding response. In brief, during the learning step the algorithm attempts to reduce the global variability of the system response for each input sequence over the time. In this scheme a solution will be considered as not feasible if the system inference is negatively affected. It should be remarked that the weight matrix can not be modified during this process since our model is oriented to compute more stable maps once the causality estimation is done.

4 Simulations and discussion

In order to validate the proposal discussed in the above section we use six previously adjusted FCM taken from the work of Grau and Nápoles [13-14]. Such maps describe the behavior of some HIV mutations related to their resistance to existing antiviral drugs. Accordingly, the authors described the HIV *protease* protein as a simple FCM where each sequence position¹ is taken as a map concept, while another node for the resistance target is also defined. Then, all the neurons are fully connected; also there exist causal links between each sequence position and the resistance concept. It means that the resistance is conditioned by the interaction of the amino acids once a specific mutation (simple or multiple) takes place, leading to different levels of resistance to the target drug. This model can not represent the external influences.

It is important to mention that each map denotes the protein behavior for a specific drug: Amprenavir (APV), Indinavir (IDV), Saquinavir (SQV), Nelfinavir (NFV), Ritonavir (RTV) and also Atazanavir (ATV). Each drug has associated a high-quality filtered datasets taken from [15] consisting in reported mutations and their resistance value. The configuration of the PSO-RSVN algorithm used as optimizer is fixed as follows: 40 particles as the population size, five variable neighborhoods ($m = 5$), 80 generations, and the allowed number of evaluations without progress is set to 20. In addition, the number of times of the FCM inference process is $T=100$.

As a first analysis the stability of the resistance node of each drug for a randomly selected mutation is measured. Figure 1, 2 and 3 show the activation value of the resistance over the time for two scenarios: the solid line represents the FCM response without any modification, whereas the dashed line denotes the FCM response using the family of sigmoid functions found by the learning method. From these simulations it is possible to conclude that our proposal induces better stability features over the drug resistance target. In this case, only the resistance node was monitored since it is the decision concept, allowing to predict whether a new mutation will be susceptible to the drug or not. Note that the system response changes for next drugs: IDV, RTV and ATV. In such cases the final classification rate does not suffer any change since the resistance target for a drug is measured in a certain range instead of using a single value. However, we noticed that some FCM achieve better accuracy.

For better understanding of this issue let us analyze the behavior of the selected mutation “FKLDVFMIIIVSVTVNML” for the map IDV. This sequence has high level of resistance for the drug IDV, which means that the higher the activation value of the resistance node, the better accuracy should report the model for this instance. In fact, in the figure 1a) the FCM inference process is able to compute higher resistance after applying the learning algorithm. As a partial conclusion, four behaviors from results may be observed: (1a-1b) it possible to compute better stability from stable maps, (2a-2b) we can obtain stable maps from maps exhibiting cyclic patterns, (3a) it is possible to compute more stable maps from chaotic systems, and (3b) we can obtain a map having stable features from a FCM exhibiting a chaotic behavior.

¹ The protease sequence is defined by 99 amino acids where each position can be represented by its normalized contact energy [16]. In order to reduce the number of nodes in map, the authors use a subset of amino acids previously associated with resistance.

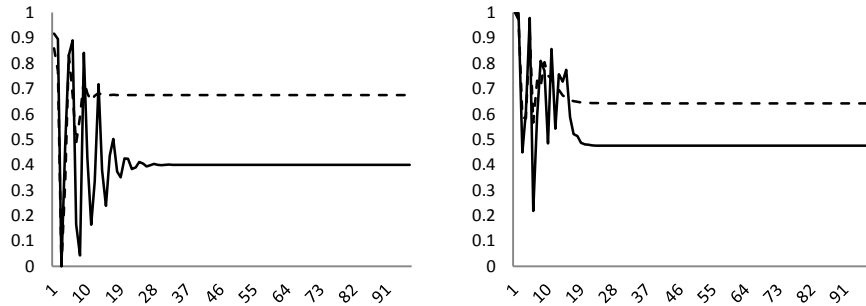


Fig. 1. Activation value of the resistance concept for a) drug IDV b) drug RTV. The *solid* line denotes the FCM response using the same function for all the neurons, whereas the *dashed* line represents the FCM output using the family of functions found by the learning scheme.

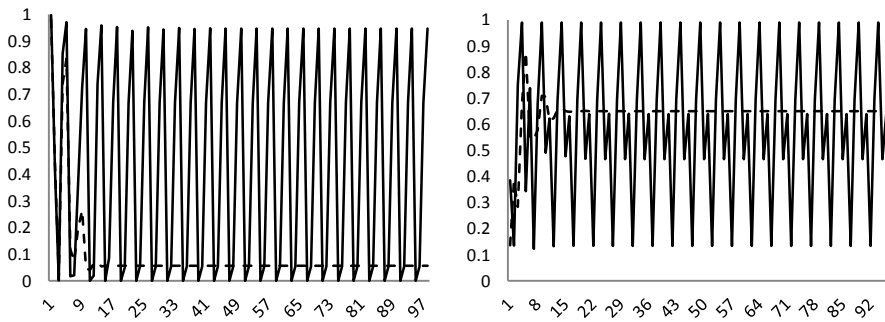


Fig. 2. Activation value of the resistance concept for a) drug ATV b) drug APV. The *solid* line denotes the FCM response using the same function for all the neurons, whereas the *dashed* line represents the FCM output using the family of functions found by the learning scheme.

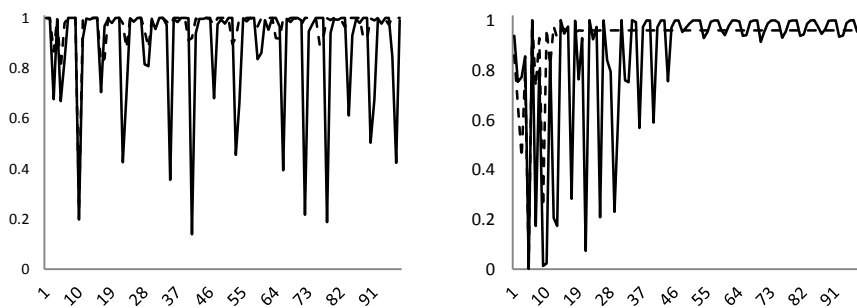


Fig. 3. Activation value of the resistance concept for a) drug SQV b) drug NFV. The *solid* line denotes the FCM response using the same function for all the neurons, whereas the *dashed* line represents the FCM output using the family of functions found by the learning scheme.

In order to generalize these results we introduce a second experiment consisting on computing the variability of the system response over the time, taking into account all historical data for each drug. In practice, this simulation is equivalent to compute the objective function (3) for each instance before applying the learning algorithm, and then computing the same formula once the learning process is done. Due to the stochastic features of the proposed learning scheme in each case we select the best solution from 10 independent simulations. Next table 1 shows the number of instances (mutations), as well as the average, mean and standard deviation for the six studied drugs with respect to the objective function. In summary, we can conclude that proper selection of the family of sigmoid functions lead to better stability.

Table 1. Variability of the system response over the time for all instances (mutations).

Drug	Mutations	Before			After		
		Average	Median	Std. deviation	Average	Median	Std. deviation
APV	96	39.09962	39.83242	1.57105	1.98303	1.96438	0.34534
ATV	69	13.84345	13.77815	2.18643	2.43406	1.87842	1.03477
IDV	137	10.58505	10.38378	4.25156	3.78721	3.53374	0.97868
RTV	151	19.19583	19.26046	0.75602	7.52529	7.37976	0.70137
NFV	204	19.59037	32.66005	14.4047	2.835233	2.78300	0.29764
SQV	139	13.92619	12.02521	6.83530	6.158739	6.30550	0.56349

Why it is desirable more stable systems? To answer this question let us analyze the inference process for the map SQV regarding the selected mutation “FKLDVFMIGV PVISTVNML”. It has a high level of resistance to drug SQV. When the same treshold function is used for all the nodes, the activation level of the decision node has lower degree of resistance towards the end, and hence the sequence may be erroneously classified as susceptible. However, using the family of sigmoid functions found by the learning algorithm the final neuron is more stable, although the biological system remains cahotic. In a few words, after aplying the methodology discussed here, the map SQV will be more chance to correctly classify new mutations.

5 Conclusions

Fuzzy Cognitive Maps are a useful neurofuzzy technique for modeling and simulation which have been successfully applied to numerous real world problems. However, they are regularly considered as closed systems, where the effects of external factors over neurons are omitted. Besides, most of the existing learning algorithm are mainly oriented to the transforming the causal weight matrix, ignoring other aspects such as the system convergence. Inspired on these limitations, the present paper proposed a new learning approach for estimating the proper threshold function for each neuron in sigmoid FCM. It attempts to efficiently simulate the effects of external stimulus over the concepts, where the stability of the modeled system is the main goal. By doing so, we used a Swarm Intelligence based approach with diversity control, for computing better estimations during the optimization of the objective function.

In order to validate our proposal we used six FCM concerning the field of the HIV protein analysis. From these results we can definitely conclude that, after applying the learning methodology, adjusted FCM exhibit more stability. In addition, we observed that (i) it is possible to achieve better stability from stable maps, (ii) it is possible to compute stable maps from maps exhibiting cyclic patterns, (iii) it is possible to obtain more stable maps from chaotic systems, and finally (iv) it is possible to accomplish a stable map from a FCM exhibiting a chaotic behavior. In this sense more stable maps allows to extract more consistent patterns from the system behavior. The future work will be focused on studying the FCM convergence, but now from the point of view of the causal links characterizing the neurons interaction.

References

1. Kosko, B.: Fuzzy Cognitive Maps. *Int. Journal of Man-Machine Studies*. 24, 65--75 (1986)
2. Kosko, B.: *Neural Networks and Fuzzy systems, a dynamic system approach to machine intelligence*. Prentice-Hall, Englewood Cliffs (1992)
3. Papageorgiou, E.P.: *Learning Algorithms for Fuzzy Cognitive Maps—A Review Study*. *IEEE Tran. on Systems, Man, and Cybernetics*. 42, 150--163 (2012)
4. Papageorgiou, E.P.: *Review study on Fuzzy Cognitive Maps and their applications during the last decade*. In: *IEEE International Conference on Fuzzy Systems*, pp. 828--835 (2011)
5. Bueno, S., Salmeron, J.L.: *Benchmarking main activation functions in Fuzzy cognitive maps*. *Expert Syst. Appl.* 36, 5221--5229 (2009)
6. Kosko, B.: *Hidden patterns in combined and adaptive knowledge networks*. *International Journal of Aproximate Reasoning*. 2, 377--393 (1988)
7. Tsadiras, A.K.: *Comparing the inference capabilities of binary, trivalent and sigmoid fuzzy cognitive maps*. *Information Science*. 178, 3880--3894 (2008)
8. Stylios, C.D., Groumpos, P.P.: *Mathematical Formulation of Fuzzy Cognitive Maps*. In: *7th Mediterranean Conference on Control and Automation*, pp. 28--30 (1999)
9. Kennedy, J., Russell, C.E.: *Swarm Intelligence*. Morgan Kaufmann Publishers (2001)
10. Kennedy, J., Eberhart, R.: *Particle Swarm Optimization*. In: *IEEE International Conference on Neural Networks*, pp. 1942--1948. Australia (1995)
11. Nápoles, G., Grau, I., Bello, R.: *Particle Swarm Optimization with Random Sampling in Variable Neighbourhoods for solving Global Minimization Problems*. In: Dorigo, M., et al. (eds.) *ANTS 2012. LNCS*, vol. 7461, pp. 352--353. Springer, Heidelberg (2012)
12. Nápoles, G., Grau, I., Bello, R.: *Constricted Particle Swarm Optimization based Algorithm for Global Optimization*. *POLIBITS - Research journal on Computer science and computer engineering with applications*. 46, 5--12 (2012)
13. Grau, I., Nápoles, G., León, M., Grau, R.: *Fuzzy Cognitive Maps for Modelling, Predicting and Interpreting HIV Drug Resistance*. In: J. Pavón et al. (eds.) *IBERAMIA 2012. LNAI*, vol. 7637, pp. 31--40. Springer, Heidelberg (2012)
14. Nápoles, G., Grau, I., León, M., Grau, R.: *Modelling, aggregation and simulation of a dynamic biological system through Fuzzy Cognitive Maps*. In: Batyrshin, I., et al. (eds.) *MICAI 2012. LNCS*, vol. 7630, pp. 188--199. Springer, Heidelberg (2013)
15. Stanford HIV Drug Resistance Database, <http://hivdb.stanford.edu>
16. Miyazawa, S., Jernigan, R.L.: *Contacts energies Self-Consistent Estimation of Inter-Residue Protein Contact Energies Based on an Equilibrium Mixture Approximation of Residues*. *PROTEINS: Structure, Function, and Genetics*. 34, 49--68 (1999)