The Cuban Scientist

Year 2021 Volume 2 | Issue 2

In this issue:

Cuban Sci. 2021, 2(1): 1–2 — Natural Sciences
Bidirectional Pooling for Deep Neural Networks
Marilyn Bello, Gonzalo Nápoles, Koen Vanhoof, Rafael Bello

Cuban Sci. 2021, 2(1): 3–4 — Natural Sciences
Dual phase magnetic functionalization of multicomponent alloys
A. Quintana-Nedelcos

Cuban Sci. 2021, 2(1): 5–6 — Natural Sciences
Theoretical Analysis of the State Space of Fuzzy Cognitive Maps using Shrink Functions
Leonardo Concepción, Gonzalo Nápoles, Rafael Bello, Koen Vanhoof

Cuban Sci. 2021, 2(1): 7–8 — Natural Sciences
A Note on Deterministic Learning of Hybrid Fuzzy Cognitive Maps and Network Reduction Approaches
Gonzalo Nápoles, Agnieszka Jastrzebska, Carlos Mosquera, Koen Vanhoof, Władysław Homenda

Bidirectional Pooling for Deep Neural Networks

Marilyn Bello[a1,2], Gonzalo Nápoles[3], Koen Vanhoof[2], and Rafael Bello[1]

[1]Department of Computer Science, Central University of Las Villas, Cuba
[2]Faculty of Business Economics, Universiteit Hasselt, Belgium
[3]Department of Cognitive Science & Artificial Intelligence, Tilburg University, The Netherlands

This work introduces a new neural network architecture that uses bidirectional associations-based pooling to extract high-level features and labels from multi-label data. Unlike the pooling approaches reported in the literature, our proposal does not require input data to have any topological properties as typically occurs with images and videos. The numerical results show that our bidirectional pooling helps reduce the number of problem features and labels while preserving the discriminatory power of the network.

Introduction

Pooling layers [1] help reduce redundancy and the number of parameters before building a multilayer (or recurrent) neural network that performs the remaining operations. Although these operators are able to deal with both single-label and multi-label classification problems (MLC) [2, 3], they are specifically aimed at reducing feature space. In the case of multi-label data, this should also be done in the label space. Despite their success, existing pooling operators [4] are focused on data with a well-defined structure (such as image and video) where the term *feature neighborhood* makes sense. However, while it is interesting to recognize faces or classify objects in images and videos, the truth is that there are other domains in which the data do not have a topological organization. In those cases, using standard pooling operators might have little sense, even when the problem at hand could benefit significantly from a deep learning solution.

The proposed network architecture

In [5], we proposed a bidirectional network composed of stacked association-based pooling layers to extract high-level features and labels in MLC problems with no specific topological organization. Unlike the classic use of pooling, this approach does not perform pooling over pixels but problem features or labels. The first pooling layer is composed of neurons denoting the problem features and labels (i.e., low-level features and labels). In contrast, neurons denote high-level features and labels extracted during the construction process in deeper pooling layers. Each pooling layer uses a function that detects pairs of highly associated neurons while performing an aggregation operation to derive the pooled neurons. Such neurons are obtained from neurons belonging to the previous layer to fulfill a certain association threshold. This model uses Pearson's correlation to estimate the association degree between two neurons. Overall, we compute the correlation matrix among features and labels and derive the degree of association of the pooled neurons from the degree of association between each pair of neurons in the previous layer. The pooling process is repeated over aggregated features and labels until a maximum number of pooling layers is reached.

Once the high-level features and labels have been extracted using the pooling operators, they are connected with one or several hidden processing layers. Finally, a decoding process [6] is performed to connect the high-level labels to the original ones by means of one or more hidden processing layers. Figure 1 depicts an example of this network architecture resulting in five high-level neurons that emerge from the association-based pooling layers. Hidden neurons in these hidden layers are equipped with can use any transfer function such as ReLU, sigmoid or hyperbolic tangent.

Numerical simulations

The performance of our model is evaluated using several MLC problems. Overall, we study how the model performs in terms of accuracy and number of features and high-level labels. Table 1 reports the number of high-level features (#HLF), feature reduction percentage (%Red-F), the number of high-level labels (#HLL), label reduction percentage (%Red-L), the accuracy obtained by the network using the extracted features and labels, the accuracy using the original features and labels (baseline model), and the loss of accuracy with respect to the baseline model.

Table 1: Performance assessment of the bidirectional association-based pooling approach.

Dataset	#HLF	%Red-F	#HLL	%Red-L	Accuracy	Baseline	Loss
D1	43	40.28%	6	0%	0.815	0.823	-0.008
D2	24	91.84%	6	0%	0.913	0.915	-0.002
D3	44	57.28%	13	7.14%	0.798	0.80	-0.002
D4	17	96.85%	22	87.43%	0.988	0.987	0.001
D5	19	96.75%	29	87.22%	0.99	0.99	0
D6	19	96.71%	50	87.5%	0.995	0.995	0
D7	20	96.85%	35	87.23%	0.991	0.991	0
D8	14	96.82%	8	0%	0.915	0.918	-0.003
D9	53	87.95%	4	0%	0.837	0.866	-0.029
D10	49	88.86%	5	16.67%	0.805	0.794	0.011
D11	4	96.67%	7	93.07%	0.965	0.965	0
D12	78	96.37%	10	95.19%	0.99	0.99	0
D13	80	92.01%	14	50%	0.928	0.988	-0.06
D14	18	96.4%	3	96.3%	0.977	0.977	0
D15	9	92.97%	3	96.3%	0.977	0.977	0

From these results, we can observe that our proposal significantly reduces the number of features and labels

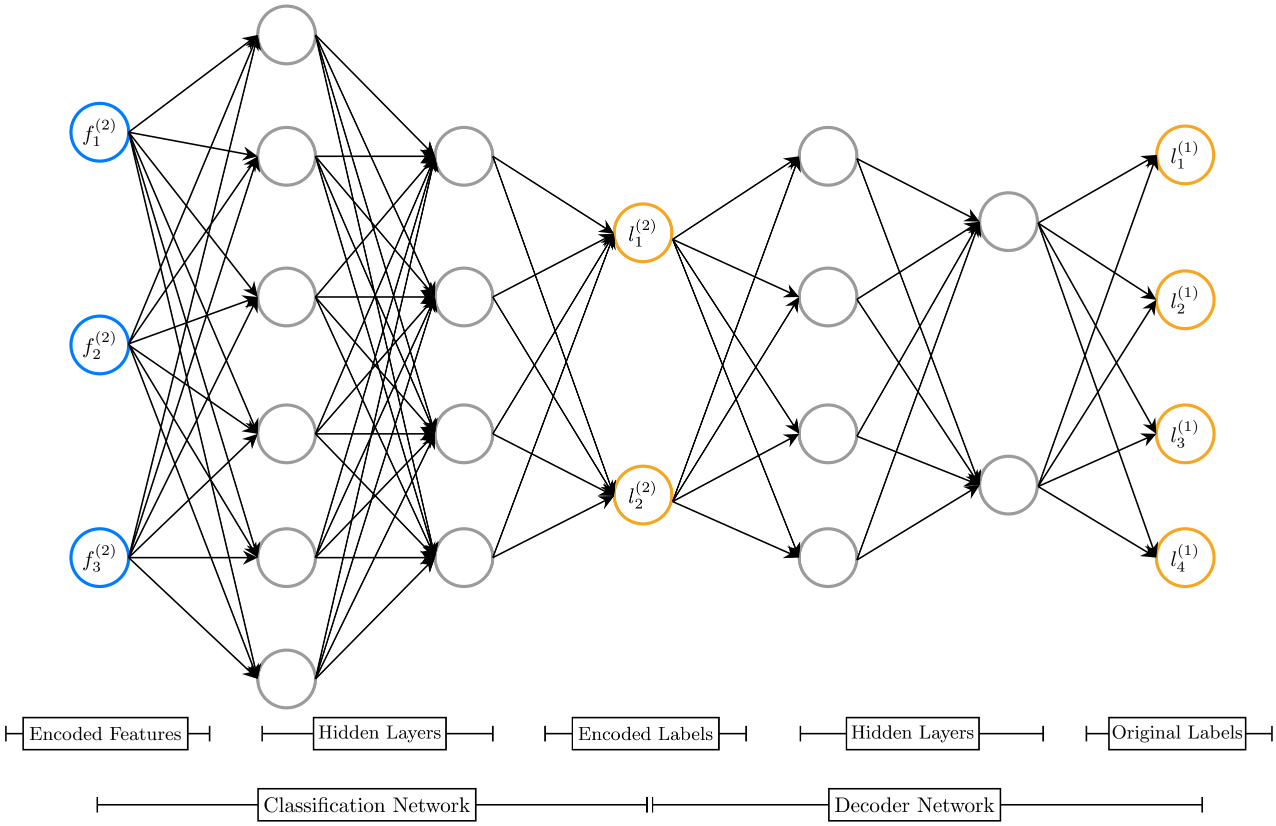

Figure 1: Neural network architecture involving three high-level features (resulting from the feature pooling step), two-high-level labels (resulting from the label feature step), four low-level labels and four hidden layers.

with a percentage reduction up to 96% and 87%, respectively. It is worth mentioning that the bidirectional association-based pooling reports a maximal accuracy loss of 0.06 for the $D13$ dataset. However, in some cases, we observed a small increase in the accuracy (e.g., dataset $D10$) even when our network was not conceived to increase the prediction rates but to obtain the same performance with smaller networks. Our proposal has no loss in accuracy for those problems having low variability in accuracy (i.e., datasets $D5 - D7$, $D11 - D12$, $D14 - D15$).

Conclusions

The numerical simulations have shown that our proposal is able to significantly reduce the number of parameters in deep feed-forward neural networks without harming their discriminatory power. Extracting high-level features and labels increases the possibility of building networks with more transparent inference models. For example, by using *post-hoc* interpretability techniques, we could shed light on the inner reasoning of the model when operating with high-level features. These techniques regularly have exponential algorithmic complexity, thus having networks with fewer parameters certainly helps reach this goal.

Notes

a. Email: mbgarcia@uclv.cu

References

[1] Goodfellow, I., Bengio, Y., Courville, A., Bengio, Y., *Deep learning,* MIT press Cambridge (2016)

[2] Charte, F., del Jesus, M.J., Rivera, A.J., *Multilabel classification: problem analysis, metrics and techniques,* Springer (2016)

[3] Gibaja, E., Ventura, S., Multi-label learning: a review of the state of the art and ongoing research, *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery,* **4** (2014) 411–444

[4] Yu, D., Wang, H., Chen, P. and Wei, Z., Mixed pooling for convolutional neural networks, *Lecture Notes in Computer Science,* **8818** (2014) 364–375

[5] Bello, M., Nápoles, G., Sánchez, R., Bello, R., Vanhoof, K., Deep neural network to extract high-level features and labels in multi-label classification problems, *Neurocomputing,* **413** (2020) 259–270

[6] Hinton, G.E., Salakhutdinov, R.R., Reducing the dimensionality of data with neural networks, *Science,* **313** (2006) 504–507

Dual phase magnetic functionalization of multicomponent alloys

A. Quintana-Nedelcos[a1]

[1]New Model Institute for Technology and Engineering (NMITE), Blackfriars St, HR4 9HS, Hereford, UK

This work focuses on the collaborative behavior of strong ferromagnetic FeCr *nanoparticles embedded in a* AlNiCo *weak magnetic matrix.*

"The whole is greater than the sum of the parts." The quote reflects on the mathematical observation, of the original phrase attributed to Ἀριστοτέλης, that, in a broader and more modern Systems Engineering language, can be rephrased as "The System is something beside, and not the same as, its elements." [1].

Illustrations of both readings can be found in modern applied sciences. For example, the intensity resulting from the interference of two beams of equal intensity can go from none to up to four times the brightness of the individual beam. Another example: ferromagnetic Heusler alloys are made of non-ferromagnetic elements such as Manganese, Nickel and Tin [2].

In this contribution, we investigate the increased magnetization of NiCoFeCrAl$_x$ multicomponent alloy by means of its FeCr-NiCoAl$_x$ dual phase functionalization.

For this study, bulk samples of stoichiometric composition Ni$_{0.5}$CoFeCr$_{0.5}$-Al$_x$ (named Al0.0, Al1.0 and Al1.5 for $x = 0.0$, 1.0 and 1.5, respectively) were arc-melted at least three times before being heat-treated at 1423 K for 10 h in Argon atmosphere [3, 4].

Figure 1 shows the magnetization vs. temperature measurements for applied fields of 25 mT (a) and 1.0 T (b). Notice the evolution from a single characteristic Curie Temperature (T_C) for the Al0.0 sample, to a double-magnetic transition with Al addition. The split of the magnetic transition is found to be concomitant to the transition from a single FCC (Al0.0) transition into a dual BCC/B2 phase (Al1.0 and Al1.5) [3]. Also notice the formation and segregation of FeCr-nanoparticles (NPs) within an AlNiCo-rich matrix with Al addition (Figure 2a, 2b and 2c). The two distinctive values of T_C for the fully segregated Al1.0 and Al1.5 samples correspond to ${T_\mathrm{C}}^{\mathrm{FeCr}} > 850\,\mathrm{K}$ and ${T_\mathrm{C}}^{\mathrm{AlNiCo}} < 700\,\mathrm{K}$, respectively [3].

FeCr-NPs were found to make the greatest contribution to the total magnetisation in the Al1.0 and Al1.5 samples. However, the jump in magnetisation occurring at ${T_\mathrm{C}}^{\mathrm{AlNiCo}}$ is approx. 5 times larger than the magnetisation change observed at ${T_\mathrm{C}}^{\mathrm{FeCr}}$. (The result can seem counterintuitive at first sight.) In turn, the saturation magnetisation (M_s) at 300 K and 1.0 T is 100 Am^{2}kg^{-1} and 70 Am^{2}kg^{-1} for samples Al1.0 and Al1.5, respectively, which is higher than the 62 Am^{2}kg^{-1} measured for sample Al0.0 at its saturation ($T = 50\,\mathrm{K}$) and only 20 Am^{2}kg^{-1} at 300 K (Figure 1).

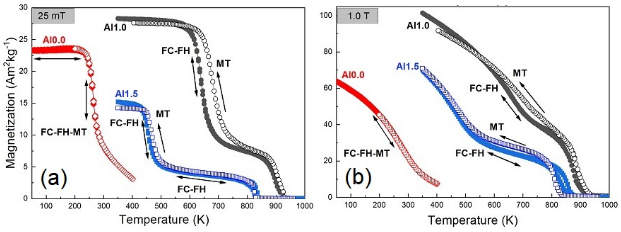

Figure 1: Shows the FC-FH path of direct M(T) measurements curves (solid symbols) and the indirect magnetization dependence on temperature (MT) curve obtained from isothermal measurements of the virgin loop in a decreasing temperature path (open symbols) for samples Al0.0 (red), Al1.0 (black) and Al1.5 (blue) at applied fields of 25 mT (a), and 1.0 T (b).

Thus, with the addition of the paramagnetic Al, the magnetic properties of the Al0.0 sample are considerably enhanced, enlarging the FM region (i.e. shifting T_C for more than 170 K towards higher temperatures) and increasing M_s by 61% and 16% for Al1.0 and Al1.5, respectively.

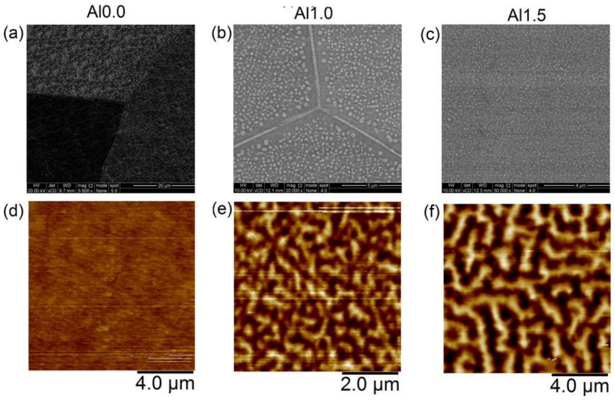

Figure 2: SEM (a, b, and c) and MFM (d, e, and f) image of Al0.0 (a and d), Al1.0 (b and e), and AL1.5 (c and f) at different magnification.

In order to explain the findings, a phenomenological approach was used where the collaborative interaction amongst NPs was taken into account. The following conclusions were reached:

1. FM-NPs of FeCr rich phase forms at ${T_\mathrm{C}}^{\mathrm{FeCr}}$.
2. As the temperature decreases, with the decrease of the thermal energy in the region $LTT < T < HTT$, dipolar interaction among NPs is favoured

for particles of larger size (i.e. Al1.0 sample). Unfavoured NPs' interaction in the Al1.5 sample show an SPM-like collective behaviour where NPs are weakly linked to their neighbours.

3. As the sample reaches ${T_C}^{\mathrm{AlNiCo}}$, the matrix becomes FM and provides the conditions to form the exchange-bridge that will enhance the FM-NPs interaction. As a consequence, already strong linked NPs of the Al1.0 sample increase the FM-like short-range interactions, which, in turn, increases the long-term dipolar repulsion that forms a highly fragmented AFM-like domain structure to minimise the energy of the system (Figure 2e). On the other hand, the smaller NPs of the Al1.5 sample are weakly ferromagnetically aligned, which, in turn, decreases the long-range dipolar repulsion interactions and allows for a wider stripes labyrinth domain structure to form (Figure 2f).

Notes

a. Email: ris.quintana-nedelcos@nmite.ac.uk

References

[1] `https://se-scholar.com/se-blog/2017/6/23/who-said-the-whole-is-greater-than-the-sum-of-the-parts`

[2] Quintana-Nedelcos, A., Llamazares, J.S., Daniel-Perez, G., Enhanced magnetocaloric effect tuning efficiency in Ni-Mn-Sn, *Journal of Magnetism and Magnetic Materials,* 441, 188-192

[3] Quintana-Nedelcos, A., Leong, Z., Morley, N.A., Study of dual-phase functionalisation of NiCoFeCr-Alx multicomponent alloys for the enhancement of magnetic properties and magneto-caloric effect, *Materials Today Energy,* 20, 100621

[4] Morley, N.A., Lim, B., Xi, J., Quintana-Nedelcos, A., Leong, Z., Magnetic properties of the complex concentrated alloy system $\mathrm{CoFeNi_{0.5}Cr_{0.5}Al}_x$, *Scientific reports,* 10(1), 1-12

Theoretical Analysis of the State Space of Fuzzy Cognitive Maps using Shrink Functions

Leonardo Concepción[a] [1,2], Gonzalo Nápoles[3], Rafael Bello[1], and Koen Vanhoof[2]

[1]Department of Computer Science, Universidad Central de Las Villas, Cuba
[2]Faculty of Business Economics, Hasselt Universiteit, Belgium
[3]Department of Cognitive Science & Artificial Intelligence, Tilburg University, The Netherlands

We proposed definitions and theorems regarding Fuzzy Cognitive Maps (FCMs), which allow estimating bounds for the activation value of each neuron and analyzing the covering and proximity of feasible activation spaces. The main theoretical findings suggest that the state space of any FCM model equipped with transfer F-functions shrinks infinitely with no guarantee for the FCM to converge to a fixed point but to its limit state space. This result, in conjunction with the covering and proximity values of FCM-based models, helps to understand their poor performance when solving complex simulation problems.

Introduction

Fuzzy Cognitive Maps (FCMs) [1] are recurrent neural networks for modeling complex systems. Existing theoretical studies on FCMs are mainly devoted to convergence issues, commonly covering the existence and uniqueness of fixed points [2, 3]. Other results reported in [4, 5, 6] address the convergence of FCM models used in prediction/classification scenarios.

Concerning the theoretical analysis of FCMs' dynamics, we summarize our paper *Unveiling the Dynamic Behavior of Fuzzy Cognitive Maps* [7]. First, we introduce several definitions and theorems that allow studying the dynamic behavior of FCMs equipped with monotonically increasing functions bounded into non-negative intervals. The strong version of our theorem proves that the state space of an FCM shrinks infinitely and converges to a so-called *limit state space*. This allows envisaging the FCM model's behavior before the inference stage. As a second contribution, we explore the covering and proximity of *feasible activation spaces*, which help explain why FCMs sometimes perform poorly when solving complex prediction problems. In other words, we should not expect impressive prediction rates when the model has low covering values, as the FCM feasible state space is small.

Shrink Functions and State Space Estimation in FCM-based Models

We define F as the set of all monotonically increasing functions bounded into non-negative intervals. Also, let $f_i \in F$ be the transfer function used in the activation process of neuron C_i in the FCM. In [7], we refer to an F-function as any function belonging to F.

Let $\mathcal{H}_W$ and $\mathcal{H}_T$ be functions that take an FCM-based model $\mathcal{M}$ and a feasible state space at the t-th iteration $\mathcal{S}^{(t)}$ for this map and return a feasible state space at the $(t{+}1)$-th iteration $\mathcal{S}^{(t+1)}$ for the same map. While $\mathcal{H}_W$ uses the weight matrix W of $\mathcal{M}$ to calculate a feasible state space for the $(t+1)$-th iteration, $\mathcal{H}_T$ uses the FCM's topology only. Based upon estimated bounds for the successive activation values and from the monotonically increasing property of $f_i \in F$, we assert that over the same FCM, these two shrink functions transform feasible state spaces into state spaces which are also feasible.

To show that FCMs are not completely unpredictable, we propose two theorems as the pillars of our state-space estimation: the *Weak Shrinking State Space* and the *Strong Shrinking State Space*. The former asserts that the state spaces shrink from one iteration to the next one, although it is possible that $\mathcal{S}^{(t)} = \mathcal{S}^{(t+1)}$, which would imply that $\mathcal{S}^{(t)} = \mathcal{S}^{(t+k)}\ \forall k \in \mathbb{N}$. So, the state spaces may not shrink forever. The latter only varies in the sense that transfer functions are now bounded into open intervals. This means that the state space bounds are never reachable and hence, the state spaces will shrink forever and they will have a limit. The *limit state space* of $\mathcal{M}$ is $\mathcal{S}^{(\infty)} = \lim_{t\to\infty} \mathcal{S}^{(t)}$, when state spaces are iteratively calculated using either shrink function $\mathcal{H}_T$ or $\mathcal{H}_W$. According to simulations, $\mathcal{S}^{(\infty)}$ often contains a single point.

Covering and Proximity of FCM Models

In this section, we discuss two evaluation measures that help understand the properties of FCM-based systems. The *covering* quantifies the proportion of the induced activation space that is reachable by the neuron's activation values and the *proximity* measures the mean relative distance of neuron's activation values to the feasible activation spaces.

Small covering values are evidence of the reduced representativeness of induced activation space, but sometimes we desire high covering values to represent the most diverse sets of outputs. As illustrated, such measures have a straightforward connection with the *Strong Shrinking State Space Theorem.* More importantly, they help explain why FCMs sometimes perform poorly when applied to prediction problems that demand high accuracy.

Experimental Scenarios

For experimentation purposes, we generated 400 FCM-based models (200 stable and 200 unstable) with varied properties according to the number of neurons (5 to 30), weights ($[-1, 1]$ interval) and connectivity or percentage of relationships (10%, 20%, ..., 100%). The simulations reported more valuable results in the presence of stable FCM models and when the knowledge comprised into the weight set is available. Therefore, Figures 1, 2 and 3 correspond to this situation.

Figure 1 depicts the covering values resulting for this scenario. Higher connectivity values and higher number of map neurons have a considerable influence on attaining higher covering values.

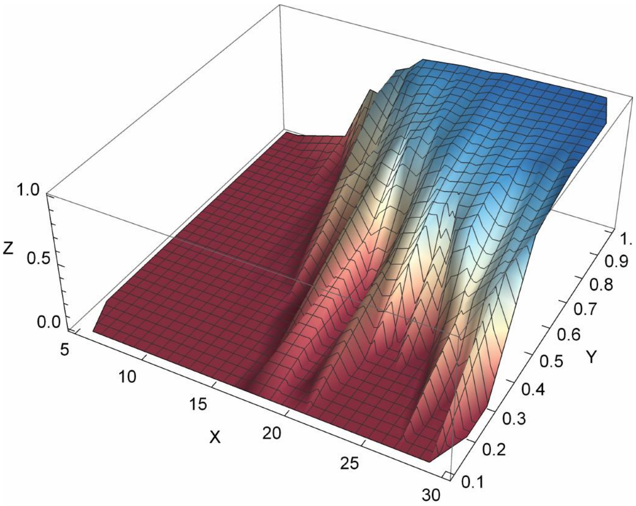

Figure 1: Axes X, Y and Z respectively refer to the number of neurons, connectivity and covering values.

A zero covering value is a computational evidence of fixed-point attractors for every one of these FCM models. According to Figure 2, at least 81 FCM model will always converge to a fixed-point attractor regardless of the initial stimulus. Moreover, in Figure 3 we can observe that nearly half of the proximity values are exactly zero and then, almost surely, 90 FCM models converge to a fixed-point attractor.

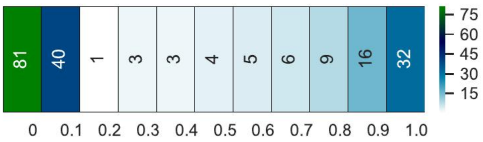

Figure 2: Distribution of covering values.

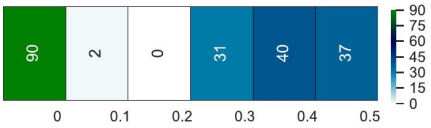

Figure 3: Distribution of proximity values.

Concluding Remarks

Our research in [7] goes a step beyond the study of fixed-point attractors, since we analyze the dynamical behavior of FCM-based models from the perspective of their state spaces. The *Strong Shrinking State Space Theorem* enunciated in this paper ensures that the feasible state space of the targeted FCMs shrinks infinitely, yet the system converges to its limit state space. As shown in the experiments, approximating an FCM's limit state space is useful to predict fixed-point attractors. Likewise, we illustrated that the covering of feasible activation spaces is often poor and irregular for FCMs with reduced network topologies. This knowledge could be injected into the learning procedure in order to improve network's performance.

Notes

a. Email: lcperez@uclv.cu

References

[1] Kosko, B. (1986). Fuzzy Cognitive Maps, *International Journal Man-Machine Studies,* **24** 1 65–75

[2] Boutalis, Y., Kottas, T., & Christodoulou, M. (2009). Adaptive Estimation of Fuzzy Cognitive Maps with Proven Stability and Parameter Convergence, *IEEE Transactions Fuzzy Systems,* **17** 4 874–889

[3] Harmati, I., Hatwágner, M., & Kóczy, L. (2018). On the Existence and Uniqueness of Fixed Points of Fuzzy Cognitive Maps, *Information Processing and Management of Uncertainty in Knowledge-Based Systems. Theory and Foundations,* 490–500

[4] Nápoles, G., Bello, R., & Vanhoof, K. (2014). How to Improve the Convergence on Sigmoid Fuzzy Cognitive Maps?, *Intelligent Data Analysis,* **18** 6S S77–S88

[5] Nápoles, G., Papageorgiou, E., Bello, R. & Vanhoof, K. (2016). On the convergence of sigmoid Fuzzy Cognitive Maps, *Information Sciences,* **349-350** 154–171

[6] Nápoles, G., Concepción, L., Falcon, R., Bello, R., & Vanhoof, K. (2018). On the accuracy-convergence tradeoff in sigmoid Fuzzy Cognitive Maps, *IEEE Transactions on Fuzzy Systems,* **26** 4 2479–2484

[7] Concepción, L., Nápoles, G., Falcon, R., Vanhoof, K., & Bello, R. (2021). Unveiling the Dynamic Behavior of Fuzzy Cognitive Maps, *IEEE Transactions on Fuzzy Systems,* **29** 5 1252-1261

A Note on Deterministic Learning of Hybrid Fuzzy Cognitive Maps and Network Reduction Approaches

Gonzalo Nápoles[a1,2], Agnieszka Jastrzębska[3], Carlos Mosquera[4], Koen Vanhoof[1], and Władysław Homenda[3]

[1]Faculty of Business Economics, Universiteit Hasselt, Belgium
[2]Department of Cognitive Science & Artificial Intelligence, Tilburg University, The Netherlands
[3]Faculty of Mathematics and Information Science, Warsaw University of Technology, Poland
[4]Faculty of Sciences and Bioengineering Sciences, Vrije Universiteit Brussel, Belgium

We have proposed a hybrid Fuzzy Cognitive Map-based model that incorporates static expert knowledge and information automatically inferred from the data. The paper introduces a fast and deterministic algorithm to construct such a model and a post-optimization procedure to fine-tune the resulting architecture. Simulations showed the superiority of our proposal in problems devoted to modeling complex systems.

Contribution overview

Fuzzy Cognitive Maps (FCMs) belong to the family of cognitive semantic models. They are represented as directed weighted graphs in which vertices are concepts (knowledge granules) and arcs correspond to relations between them. FCMs are used to visualize, model, and simulate the behavior of systems.

In our new paper [1] the focus is on an FCM-based model for simulating dynamic systems. In such a system, we often have several input variables that influence the values of the output variables. In this paper, a new hybrid approach for designing FCMs simulating such a system is presented. The proposed approach combines the ability to integrate expert knowledge into the map architecture with elements of automatic model learning from historical data. In practice, in the proposed model we can incorporate information about the weights between input variables given by experts, and the remaining weights are learned automatically. The use of the term “hybrid” refers in our case to the possibility of integrating expert knowledge about input variables into a model whose other parts are learned from the data. The specific research goals and our novel contributions can be summarized as follows:

- We developed a new method for constructing cognitive maps, in which we incorporate both static expert knowledge and knowledge extracted from the data in the learning procedure.
- We introduced a new very fast, deterministic way of learning model weights.
- We introduced a post-optimization approach for eliminating irrelevant weights and correcting the model to preserve the accuracy.

The new method

Figure 1 shows the proposed architecture with three inputs (x_1, x_2, x_3) and three outputs (y_1, y_2, y_3).

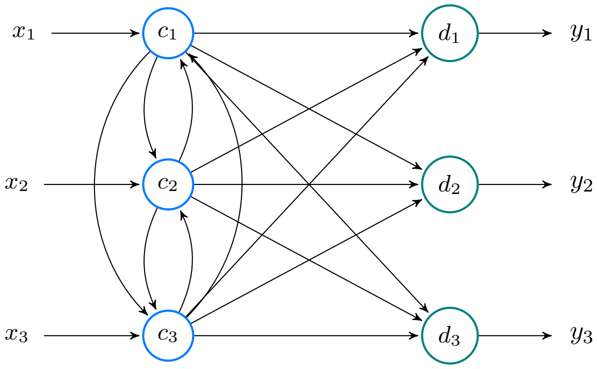

Figure 1: FCM-based model with three input neurons (c_1, c_2, c_3) and three output neurons (d_1, d_2, d_3).

The reasoning process is given in below,

$$a_{ki}^{(t+1)} = f_i\left(\sum_{j=1}^{P} w_{ji} a_{kj}^{(t)}\right), i \neq j \tag{1}$$

where P is the number of nodes in the network, k is the input activation vector index, w_{ji} is the weight connecting c_j with c_i, and $f_i(\cdot)$ is the transfer function. We adopted a simplified variant of the sigmoid function, which is depicted below,

$$f_i(x) = l_i + \frac{u_i - l_i}{(1 + e^{-\lambda_i(x-h_i)})} \tag{2}$$

such that $\lambda_i > 0$ and $h_i \in \mathbb{R}$ are parameters that define the shape of the sigmoid function.

The initial activation values $a_1^{(0)}$, $a_2^{(0)}$ and $a_3^{(0)}$ for input neurons correspond to the problem variables x_1, x_2 and x_3, respectively. Let us use N to note the number of input neurons $C = \{c_1, \ldots, c_N\}$ and M to note the number of output neurons $D = \{d_1, \ldots, d_M\}$. The weight matrix W is composed of two sub-matrices W^I and W^O. The first one contains the connections among the input neurons. W^I should ideally be defined by domain experts and it will not be modified during the learning phase. Sub-matrix W^O contains the weights between the input and the output nodes. This matrix will be learned automatically.

The inverse learning method

In the paper, we developed a learning method to compute W^O, the sub-matrix that contains the weights between the input and the output nodes.

Assume that $[X, Y]$ is a training dataset where $X = [x_{ij}], i = 1, \ldots, K, j = 1, \ldots, N$. K is the number of input instances, each instance is described with N input variables, $x_{ij} \in [0, 1]$. $Y = [y_{ij}], i = 1, \ldots, K, j = 1, \ldots, M$ is a matrix containing the true values of the M output variables for each one of the K instances.

The first step toward computing the W^O matrix is to capture the system semantics with the use of input and the weight matrix W^I. Thus, let $\Psi^{(T)}(X)$ denote an $N \times K$ matrix after performing T iterations of the FCM inference process on the input matrix X, that is $\Psi^{(T)}(X) = [a_{ij}^{(T)}], i = 1, \ldots, K, j = 1, \ldots, N$.

The second learning step computes W^O by using the proposed pseudoinverse learning rule,

$$W^O = \left(\Psi^{(T)}(X)\right)^{\ddagger} F^{-}(Y) \tag{3}$$

where $(\cdot)^{\ddagger}$ represents the Moore-Penrose (MP) inverse of a given matrix, and

$$F^{-}(Y) = \begin{bmatrix} f_1^{-}(y_{11}) & \ldots & f_i^{-}(y_{1i}) & \ldots & f_M^{-}(y_{1M}) \\ \vdots & \ddots & \ddots & \ddots & \vdots \\ f_1^{-}(y_{k1}) & \ddots & f_i^{-}(y_{ki}) & \ddots & f_M^{-}(y_{kM}) \\ \vdots & \ddots & \ddots & \ddots & \vdots \\ f_1^{-}(y_{K1}) & \ldots & f_i^{-}(y_{Ki}) & \ldots & f_M^{-}(y_{KM}) \end{bmatrix}$$

is a $K \times M$ matrix containing the inverse of the transfer functions attached to output neurons:

$$f_i^{-1}(y) = \frac{-\ln(-1-y) + h_i \lambda_i}{\lambda_i}. \tag{4}$$

We use an orthogonal projection to obtain the MP inverse. If a matrix H has linearly independent columns ($H^\top H$ is nonsingular), then $H^{\ddagger} = (H^\top H)^{-} H^\top$. In contrast, if H has linearly independent rows ($HH^\top$ is nonsingular), then $H^{\ddagger} = H^\top (H^\top H)^{-}$. The former is a left inverse because $H^{\ddagger} H = I$, while the latter is a right inverse because $HH^{\ddagger} = I$.

To overcome the issue of weights that lie outside of the desired $[-1, 1]$ interval, in the paper, we proposed a post-training weight normalization method to ensure that $w_{ji} \in [-1, 1]$. What is more, we proposed a method for superfluous weights elimination. After this step, the procedure provides the means for calibrating the weights to be retained.

Numerical simulations

In the paper, we have conducted several simulations using 35 datasets. We compared the efficiency of the new method with well-known population-based optimizers: Global-best Particle Swarm Optimization, Real-Coded Genetic Algorithm, and Differential Evolution. Figure 2 shows the box-plots associated with Mean Squared Error (MSE) for each optimization model after being tested on datasets used for comparison. The results confirmed that the MP inverse learning method is outperforms the remaining optimization approaches when it comes to the prediction error.

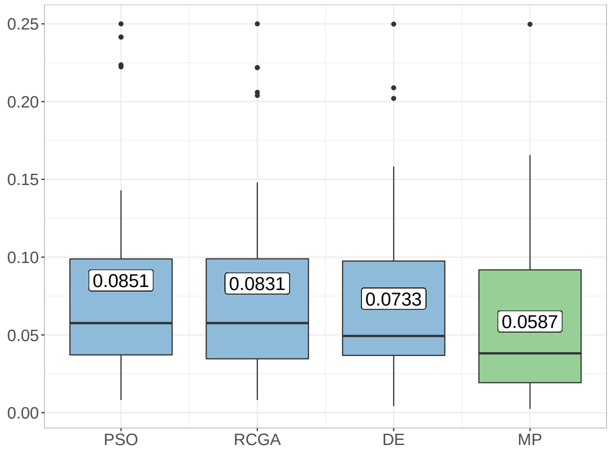

Figure 2: MSE reported by each optimization model across the 35 datasets used for simulation.

The training time of the proposed MP learning rule is significantly smaller when compared with the ones attached to state-of-the-art population-based optimizers. The average processing time of tested algorithms was as follows: Particle Swarm Optimization 4.951s, Real-Coded Genetic Algorithm 5.509s, Differential Evolution 12.964s. The new method took on average 0.003s.

Concluding remarks

The new method based on the Moore-Penrose pseudoinverse is fast and deterministic. Both those qualities are rarely seen together in the domain of FCM learning. The numerical simulations have shown that our system is able to significantly outperform state-of-the-art population-based algorithms in terms of both simulation error and training time.

*

Notes

a. Email: G.R.Napoles@tilburguniversity.edu

*

References

[1] Nápoles, G., Jastrzębska, A., Mosquera, C., Vanhoof, K., Homenda, W., Deterministic learning of hybrid Fuzzy Cognitive Maps and network reduction approaches, *Neural Networks,* **Volume 124** (2020) Pages 258–268