

Partitive Granular Cognitive Maps to Graded Multilabel Classification

Peer-reviewed author version

NAPOLES RUIZ, Gonzalo; Falcon, Rafael; PAPAGEORGIOU, Elpiniki; Bello, Rafael & VANHOOF, Koen (2016) Partitive Granular Cognitive Maps to Graded Multilabel Classification. In: 2016 IEEE International Conference on Fuzzy Systems (FUZZ-IEEE), IEEE,p. 64-71.

DOI: 10.1109/FUZZ-IEEE.2016.7737848

Handle: <http://hdl.handle.net/1942/22997>

Partitive Granular Cognitive Maps to Graded Multilabel Classification

Gonzalo Nápoles^{*†}, Rafael Falcon[‡], Elpiniki Papageorgiou^{*§}, Rafael Bello[†] and Koen Vanhoof^{*}

^{*}Hasselt Universiteit Campus Diepenbeek

Agoralaan Gebouw D, BE3590 Diepenbeek, Belgium

Email: gonzalo.napoles@student.uhasselt.be; koen.vanhoof@uhasselt.be

[†]Computer Science Department, Central University of Las Villas

Carretera Camajuaní Km 5.5, 54830 Santa Clara, Cuba

Email: gnapoles@uclv.edu.cu; rbellop@uclv.edu.cu

[‡]Electrical Engineering and Computer Science, University of Ottawa

800 King Edward Ave., Ottawa, ON K1N 6N5, Canada

Email: rfalcon@uottawa.ca

[§]Department of Computer Engineering, Technological Educational Institute of Central Greece

30 Km Old National Road Lamia-Athens, TK 35100, Lamia, Greece

Email: epapageorgiou@mail.teiste.gr

Abstract—In a multilabel classification problem, each object gets associated with multiple target labels. Graded multilabel classification (GMLC) problems go a step further in that they provide a degree of association between an object and each possible label. The goal of a GMLC model is to learn this mapping while minimizing a certain loss function. In this paper, we tackle GMLC problems from a Granular Computing perspective for the first time. The proposed schemes, termed as *partitive granular cognitive maps* (PGCMs), lean on Fuzzy Cognitive Maps (FCMs) whose input concepts represent cluster prototypes elicited via Fuzzy C-Means whereas the output concepts denote the set of existing labels. We consider three different linkages between the FCM's input and output concepts and learn the causal connections (weight matrix) through a Particle Swarm Optimizer (PSO). During the exploitation phase, the membership grades of a test object to each fuzzy cluster prototype in the PGCM are taken as the initial activation values of the recurrent network. Empirical results on 16 synthetically generated datasets show that the PGCM architecture is capable of accurately solving GMLC instances.

I. INTRODUCTION

For quite some time, *classification* has been a primordial task for the Machine Learning and Data Mining communities [1] [2]. The development of a classification model is usually oriented towards discovering the underlying mapping between a set of input objects that are characterized by a multidimensional feature space and the class label assigned to them.

More recently, *multilabel classification* [3] [4] has emerged as an extension of traditional classification problems in which each input object is associated with multiple labels; for instance, a patient and his symptoms, a document and its topics or a photograph and its themes.

Chen et. al. [5] went a step further and introduced *graded multilabel classification* (GMLC), in which an object is associated with each possible label to a certain extent/degree. The relevance of each label is no longer a binary but a fuzzy membership, typically in the interval [0;1]. From this

perspective, GMLC can be interpreted as a set of *ordinal classifications* given the ordering of the objects in the set induced by their membership grades to a label.

On the other hand, *granular computing* (GrC) [6] [7] [8] is a vibrant research discipline devoted to the design of high-level information granules and their inference frameworks. By adopting more symbolic constructs such as sets, intervals or similarity classes to describe numerical data, GrC has established a more human-centric manner of interacting with and reasoning about the real world. The emergence of *granular classifiers* [9] [10] [11] [12] [13] [14] is an important GrC manifestation that helps coping with the volume and velocity components of the V's vector characterizing Big Data [15].

In this paper, we tackle GMLC problems from a GrC angle for the first time, to the best of the authors' knowledge. Our contributions are as follows: (1) we proposed a granular system termed *granular partitive cognitive map* (PGCM) that relies upon Fuzzy Cognitive Maps (FCMs) whose input concepts represent cluster prototypes elicited via a well-known partitive clustering algorithm, viz Fuzzy C-Means (FC-MEANS) [16], whereas output concepts denote the set of problem labels; (2) we consider three FCM topologies determined by different causal connections among the sets of input and output concepts; (3) we learn the ensuing weight matrix through a Particle Swarm Optimizer (PSO); (4) we generate 16 synthetic GMLC datasets out of the UCI Machine Learning repository and (5) we evaluate the three PGCM models in presence of these datasets, with the empirical evidence confirming that the proposed architecture is capable of accurately solving GMLC instances.

The paper continues with Section II briefly reviewing relevant concepts to this study. Section III dissects the PGCM model while Section IV unveils the empirical analysis undertaken to validate the three proposed granular systems. Conclusions and future directions are found in Section V.

II. RELATED WORK

This section briefly goes over several topics of interest to the study.

A. Graded multilabel classification

GMLC [5] was formalized as an extension of multilabel classification problems [3] where one is interested in the degree of association between a data point and a label. The authors reduce their approach to both a set of ordinal classification problems and a set of multilabel classification problems so they can apply existing performance metrics therein.

Brinker et. al. [17] reformulated the original GMLC in terms of preferences between the labels and their scales. They solve this new scenario via pairwise comparisons using three different variants of Calibrated Label Ranking.

Lastra et. al. [18] introduced a nondeterministic learner based on a binary relevance strategy that returns a prediction interval whenever the classification is uncertain for a label. The authors claim that using narrow intervals for a label prediction greatly improves the classification accuracy.

All of the above approaches treat GMLC membership grades as pertaining to an ordinal scale (e.g., rate a movie using 1-5 stars according to different categories). In this paper we generalize these grades as belonging to a numerical scale. We use the activation values in $[0;1]$ generated by a FCM (see Section III) as the output of our granular model for solving a GMLC instance.

B. Fuzzy C-Means

Fuzzy C-Means (FC-MEANS) [16] is one of the most popular partitive clustering techniques. Created by Bezdek et. al., FC-MEANS is a generalization of the hard c-means method in that it defines membership grades of a data point to all c clusters, where c is a user-defined parameter. The original dataset is thus represented as a collection of cluster prototypes $P = (P_1, P_2, \dots, P_c)$ and a fuzzy partition matrix $U = [u_{ji}]$ denoting the membership grade of the i -th data point to the j -th cluster. The alternate optimization of the objective function Q in Equation (1) capturing the sum of weighted Euclidean distances between each data points X_i and cluster prototypes P_j drives the entire algorithmic process. The fuzzy partition matrix U is iteratively updated as displayed in Equation (2).

$$Q(U, P) = \sum_{j=1}^c \sum_{i=1}^N (u_{ji})^m \mathcal{D}(X_i, P_j)^2 \quad (1)$$

$$u_{ji} = \left(\sum_{q=1}^c \frac{\mathcal{D}(X_i, P_j)}{\mathcal{D}(X_i, P_q)} \right)^{-\frac{2}{m-1}} \quad (2)$$

C. Fuzzy Cognitive Maps

Fuzzy Cognitive Maps (FCMs) are recurrent neural networks [19] whose neurons and weights respectively denote the system concepts and the strength of the causal connection (confined to $[-1;1]$) for each pair of concepts [20]. The activation value of a neuron could take values in either the

$[0;1]$ or the $[-1;1]$ range. The higher the activation value of a concept, the stronger its influence over the entire system.

The update of the activation value $A_i^{(t+1)}$ of the i -th FCM concept at time $t + 1$ is reflected in Equation (3), where M is the number of FCM concepts, w_{ji} denotes the weight of the edge from concept C_j to concept C_i and $f(\cdot)$ is a monotonically non-decreasing, nonlinear function used to transform the activation value of each concept, i.e., the weighted combination of the activation levels. In this paper we will focus on sigmoid functions as the choice for $f(\cdot)$ since they exhibit superior prediction capabilities [21].

$$A_i^{(t+1)} = f \left(\sum_{j=1}^M w_{ji} A_j^{(t)} + w_{ii} A_i^{(t)} \right), i \neq j \quad (3)$$

Equation (3) is iteratively repeated until either the map converges or a maximum number of time steps T is reached. Each time step t induces a new vector with the activation of all concepts; it is expected that after multiple iterations, the map will arrive at one of the following states: (i) fixed equilibrium point, (ii) limited cycle or (iii) chaotic behavior [22]. The map is said to have converged if it reaches a fixed-point attractor. The FCM's output to the initial activation vector is the final vector of activation values after the stop criterion is reached.

D. Granular Cognitive Maps

FCMs have been augmented with different types of information granules, thus giving rise to high-level constructs known as *Granular Cognitive Maps* (GCMs). Pedrycz [23] mentions the allocation of information granularity as a pivotal driving force behind the development of these types of granular structures and describes five protocols as its realization mechanisms.

Pedrycz [24] and his collaborators [25] put forth a granular representation of time series in which FCM nodes are generated after the cluster prototypes induced by FC-MEANS over the space of amplitude and change of amplitude.

Homenda et. al. [26] adopted numeric intervals as the granulation vehicle for their GCM weight matrix. That is, each FCM weight is no longer a number but an interval. The authors elaborate on three methodologies for building a GCM from scratch by maintaining an adequate balance between specificity and generality in the design of the interval-based FCM weights and the ensuing map operations. They found the resulting GCMs had a good degree of coverage without a loss in precision.

Nápoles et. al. [12] [13] employed rough set theory [27] as the underlying information granulation vehicle in their GCMs. Their FCM input nodes are the positive, negative and boundary regions induced by a similarity relation over a subspace of the attribute set whereas the FCM output nodes are the values taken by the decision attribute. The weight matrix is learned through the Harmony Search [12] metaheuristic. The resulting granular system was named *Rough Cognitive Network* and successfully applied to intrusion detection in computer networks [13].

Our PGCM model is closer to the works in [24] and [25]; however, we are not concerned with time series prediction but with solving GMLC problems. The main differences between our approach and the previous models are related to: (1) the design of the network topology to solve GMLC problems; (2) the use of standard FCMs instead of high-order FCMs and (3) the inclusion of convergence features into the learning scheme.

III. PROPOSED METHODOLOGY

In this section we introduce a new granular fuzzy cognitive map for addressing GMLC problems. In these problems, a data point $x \in X$ may be associated with a subset of labels $\mathcal{L}_x \in \mathcal{L}$, where \mathcal{L} denotes the set of all possible labels. On the other hand, the data-point-to-label relation could be partial and thus quantified in the $[0;1]$ range. Therefore, each instance $x \in X$ belongs to a label $\mathcal{L}_i \in \mathcal{L}$ to a certain degree; this implies that \mathcal{L}_x comprises a fuzzy subset of \mathcal{L} . Hence, the goal of a GMLC solver is to learn a model $\mathcal{M} : X \rightarrow [0;1]^{|\mathcal{L}|}$ minimizing the prediction loss. However, computing this model could be challenging since this is equivalent to solving $|\mathcal{L}|$ regression problems where labels could be correlated.

The proposed granular model comprises two steps, namely, (1) the construction of the network topology and (2) the estimation of causal weights among neurons. Next we explain how to perform these steps from a training set.

A. Designing the Network Topology

The main goal of this step is to construct the topology of the granular FCM by using information granules coming from the well-known FC-MEANS partitive fuzzy clustering algorithm [16] for numerical domains. This method determines a structure of the data given a user-defined number of clusters c and returns the membership degree u_{ji} of each data point x_i to each cluster c_j , $j = 1, \dots, c$. Moreover, the algorithm associates a prototype P_j to the j -th fuzzy cluster, which is computed according to Equation (4). Such cluster prototypes define a simplified representation of the input data and have been adopted for solving challenging prediction problems [23][24][25].

$$P_j = \sum_{i=1}^N u_{ji}^m X_i / \sum_{i=1}^N u_{ji}^m \quad (4)$$

Figures 1 - 3 show the proposed GCMs where fuzzy cluster prototypes are mapped as input neurons and the problem labels constitute the set of output neurons. Each neural processing entity leans on a sigmoid transformation function to keep its activation value within the proper range. In the first topology (GCM-1), input neurons are fully connected and there exist causal connections between each input neuron (prototype) and each label. The second topology (GCM-2) is quite similar to the first one; however, output neurons are also fully connected among themselves to capture inter-label correlations. The last topology (GCM-3) extends the second model by adding connections from output to input neurons, and therefore the whole GCM becomes entirely connected.

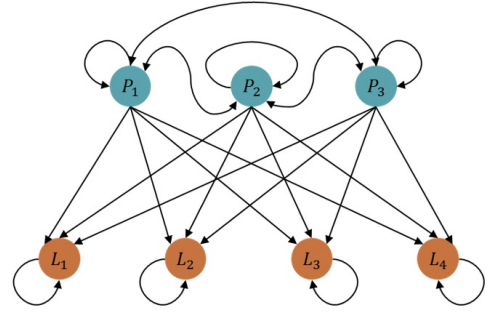


Fig. 1. Topology of the first partitive granular FCM (GCM-1)

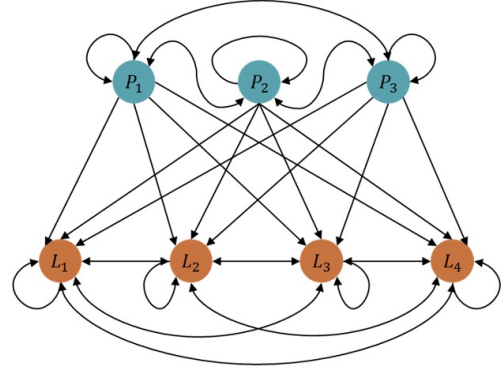


Fig. 2. Topology of the second partitive granular FCM (GCM-2)

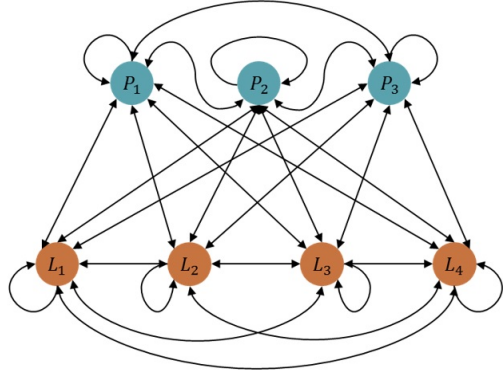


Fig. 3. Topology of the third partitive granular FCM (GCM-3)

Let us assume that c is the number of cluster prototypes and $|\mathcal{L}|$ the number of labels. All three granular architectures will have exactly c input neurons and $|\mathcal{L}|$ output neurons; however, the number of connections in each will differ. The GCM-1 model has $c^2 + c|\mathcal{L}| + |\mathcal{L}|$ causal links. In the GCM-2 model, the resultant network has $c^2 + c|\mathcal{L}| + |\mathcal{L}|^2$ causal relations while the last topology exhibits $(c + |\mathcal{L}|)^2 = c^2 + 2c|\mathcal{L}| + |\mathcal{L}|$ causal connections. Clearly, the GCM-3 model is more complex since it involves a higher number of edges.

It should be mentioned that the term “topology” refers to the protocol for connecting input-type (prototypes) and

output-type (labels) neurons. To activate the FCM, we need to compute the fuzzy membership grade on which the target object belongs to each FC-MEANS-elicited cluster prototype. Then the typical FCM reasoning process is carried out in order to activate the set of output neurons. At the end, such sigmoid neurons will comprise the activation degree of each label for the object used to initially activate the network. The reader may notice that such topologies do not include the causal weights among neurons, and therefore the reasoning process using FCM is not possible until we set the weight matrix appropriately.

B. Adjusting the Causal Weights

We now propose a learning method for estimating the causal weight matrix that defines the PGCM system. This learning process is of supervised nature and therefore relies on historical data. As the first step, each data point $x \in X$ is explicitly divided into a pair of vectors Y_x and Z_x that indicate the values of the predictive attributes and the values of the decision labels, respectively. For the sake of simplicity, in this work we assume that $Y_x \in \mathbb{R}^N$; however, categorical attributes could be considered as well. Allowing for symbolic-type attributes only implies selecting a proper clustering method, but the remaining steps of the methodology can be adopted without further changes.

In order to estimate the causal relations from data we design a learning method that minimizes the error function in Equation (5), where K is the number of training objects, $|\mathcal{L}|$ is the number of problem labels under consideration, T denotes the number of discrete time steps adopted in the FCM reasoning rule, $\omega_t = t/T$ is the relative importance of the output $A_{ki}^{(t)}$ during the recurrent reasoning stage and Z_{ki} represents the expected activation degree for the i -th label and the k -th training object.

$$\text{minimize } \Theta(W) = \sum_{k=1}^K \sum_{i=1}^{|\mathcal{L}|} \sum_{t=1}^T \frac{2\omega_t (A_{ki}^{(t)} - Z_{ki})^2}{K|\mathcal{L}|(T+1)} \quad (5)$$

The novelty of this learning scheme is given by the active consideration of convergence issues into the optimization phase. Most FCM learning algorithms only take into account the activation value of the neurons at the last discrete time step T , thus ignoring earlier activation values[28]. In Equation (5) however, all values are considered. Moreover, the adoption of the weight ω_t allows controlling the relevance of each discrete time step. In this paper we are focused on the prediction ability of the proposed model in GMLC environments, and therefore the convergence will not be analyzed. The reader can find details about these issues in [29] [30] [31].

Equation (6) formalizes how to activate the granular FCM for the k -th object. The initial activation value of the j -th input neuron is given by the fuzzy membership grade of the attribute vector Y_k to the P_j cluster prototype, with m denoting the fuzzifier parameter. The same activation strategy is used whenever we want to exploit the network, i.e., predict the set of labels associated with a new observation.

$$A_{kj}^{(0)} = \left(\sum_{l=1}^c \frac{\mathcal{D}(Y_k, P_j)}{\mathcal{D}(Y_k, P_l)} \right)^{-\frac{2}{m-1}} \quad (6)$$

The last aspect to be considered is how to optimize the error function in Equation (5). In this study we utilize a *Swarm Intelligence* approach to estimate the weights. Particle Swarm Optimization (PSO) is an effective search method for solving challenging continuous optimization problems [32]. The PSO metaheuristic involves a set of particles, known as *swarm*, which collaboratively explore the search space trying to locate promising regions [33]. Particles are interpreted as candidate solutions for the optimization problem and they are encoded as points in an S -dimensional search space. The dimension of this search space is subject to the number of parameters to be optimized. In this case, S is the number of causal connections of the target GCM architecture; hence, we have $c^2 + c|L| + |\mathcal{L}| \leq S \leq (c + |L|)^2$ as discussed in Section III-A. During the search process, each particle navigates through the search space using its own velocity, a local memory of the best position it has obtained and the knowledge of the best solution found so far within its topological neighborhood. For more details on the PSO method, the reader may consult [32].

It should be mentioned that our learning methodology is not tied to a specific optimization procedure. In fact, other population-based search methods such as *Genetic Algorithms* or *Differential Evolution* could be adopted as well. Even *gradient-based methods* may be suitable depending on the problem at hand. In general terms, the simplicity of the granular networks becomes an important advantage when adjusting the causal weights as opposed to other neural models that comprise several layers and hidden neurons.

IV. EXPERIMENTAL ANALYSIS AND DISCUSSION

The empirical analysis conducted in this Section is rather exploratory and focused on the prediction capabilities of the three proposed granular models.

A. Dataset Generation

The lack of suitable GMLC datasets is the first hindrance to be bypassed. Existing datasets for GMLC environments such as BeLa-E [34] are oriented to the ordinal prediction of labels instead of the exact degree of each label. The same drawback is observed in existing algorithms and measures to validate numerical results. The ordinal prediction of labels is a particular case of GMLC situations and is inadequate to explore the accuracy of our models.

For example, let us assume a GMLC problem having three labels where the expected degree of each label for an object x is given by $Z_{x1} = 0.8$, $Z_{x2} = 0.6$ and $Z_{x3} = 0.4$. The ordinal prediction of such labels leads to a set of suitable solutions given by Equation (7), where \hat{Z}_{x1} , \hat{Z}_{x2} , \hat{Z}_{x3} denote the predicted degree associated with the three respective labels.

$$\Omega_x = \{ \hat{Z}_{x1}, \hat{Z}_{x2}, \hat{Z}_{x3} \in [0; 1] : \hat{Z}_{x1} > \hat{Z}_{x2} > \hat{Z}_{x3} \} \quad (7)$$

To overcome the lack of GMLC datasets, we generate new problem instances¹ from standard Machine Learning datasets² in a two-step procedure. In the first step, each class label (i.e., discrete value for the class attribute) is deemed a new problem label. In the second step, the numerical degree of each class label is estimated from a model $\mathcal{M} : X \rightarrow [0, 1]^{|\mathcal{L}|}$. This model could be constructed by using a standard classifier like Random Forests [35]. More explicitly, after training the standard classifier, each object is inferred and the value of each label is taken as the probability distribution of the nominal class attribute. In this paper we adopted the Random Forests technique [35] due to its high accuracy and increasing popularity in the Machine Learning community.

B. Performance Metric and Parameter Setting

The above scheme implies that our granular networks must be capable of approximating the behavior of Random Forests but from a GMLC viewpoint. Equation (8) shows the performance metric adopted in this research to measure the overall accuracy of a GMLC learner. The Normalized Mean Squared Error (NMSE) calculates the squared difference between the expected grade Z_{ki} and the predicted grade \hat{Z}_{ki} for each testing instance. It should be remarked that, from the standpoint of the FCM's recurrent reasoning rule, the predicted grade \hat{Z}_{ki} is the activation value of the i -th output unit (class label) at time T using the k -th object to initially activate the input-type (cluster prototype) neurons.

$$NMSE = \frac{1}{K|\mathcal{L}|} \sum_{k=1}^K \sum_{i=1}^{|\mathcal{L}|} \left(\hat{Z}_{ki} - Z_{ki} \right)^2 \quad (8)$$

In the FC-MEANS clustering algorithm, the number of clusters c ranges from 2 to 10, the fuzzifier m was set to 2.0 and the number of maximum iterations equals 50. For the PSO-based learning algorithm, we adopted 40 particles as the swarm size, 120 iterations as stop criterion, the acceleration constants $c_1 = c_2 = 2.5$ and the constriction factor calculation as suggested in [36]. As a final point, the number of discrete time steps to run the FCM recurrent inference process was set to $T = 50$, which is a reasonable value to reach convergence in the proposed networks.

C. Experimental Results

Figures 4-19 summarize the NMSE metric for the following datasets: Heart, Heart-noise, Iris variants, New thyroid variants, Balance-noise, Ecoli variants, Appendicitis and Glass variants.³ On the other hand, in such datasets the number of attributes varies from 5 to 13 while the number of labels ranges from 2 to 7. It should be noted that all predictive attributes are numerical; other PGCM extensions for handling symbolic and mixed-value attributes will be formulated as a sequel of this work.

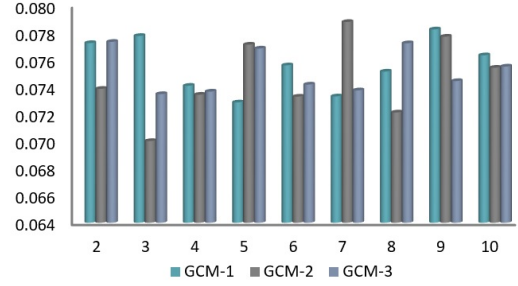


Fig. 4. NMSE as a function of the number of clusters c for the Heart dataset

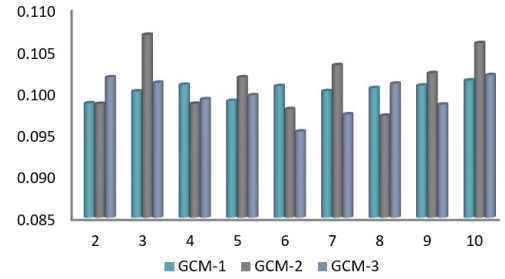


Fig. 5. NMSE as a function of the number of clusters c for the Heart-Noise dataset

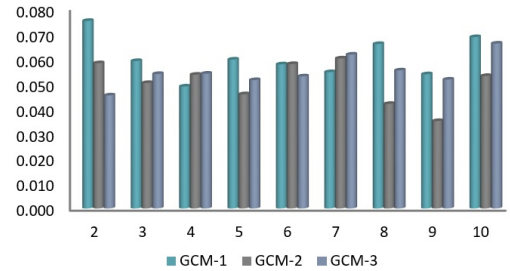


Fig. 6. NMSE as a function of the number of clusters c for the Iris dataset

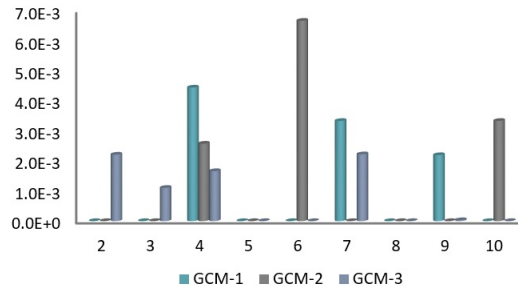


Fig. 7. NMSE as a function of the number of clusters c for the Iris0 dataset

¹Available at <http://www.eecs.uottawa.ca/~rfalc032/files/gmlc-arff.zip>

²<http://sci2s.ugr.es/keel/datasets.php>. Accessed January 7, 2016

³The notation "name#", e.g., Ecoli2, denotes that the original class distribution has been altered

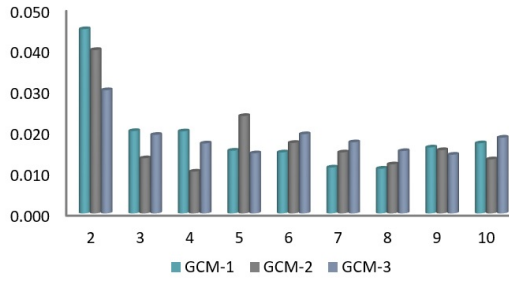


Fig. 8. NMSE as a function of the number of clusters c for the New-Thyroid1 dataset

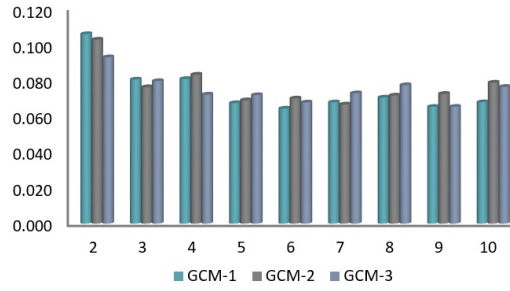


Fig. 12. NMSE as a function of the number of clusters c for the Ecoli1 dataset

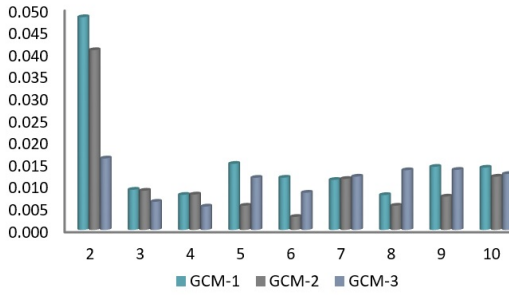


Fig. 9. NMSE as a function of the number of clusters c for the New-Thyroid2 dataset

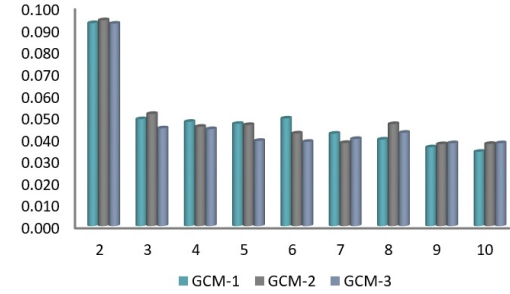


Fig. 13. NMSE as a function of the number of clusters c for the Ecoli2 dataset

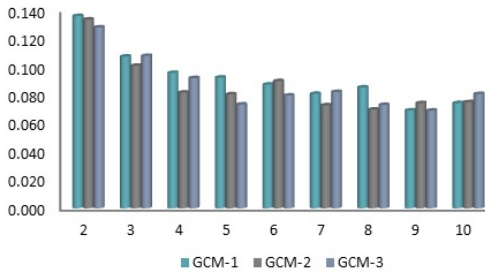


Fig. 10. NMSE as a function of the number of clusters c for the Balance-Noise dataset

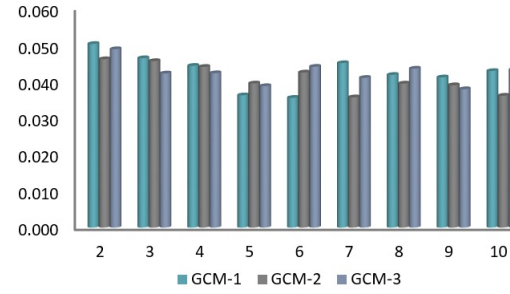


Fig. 14. NMSE as a function of the number of clusters c for the Ecoli3 dataset

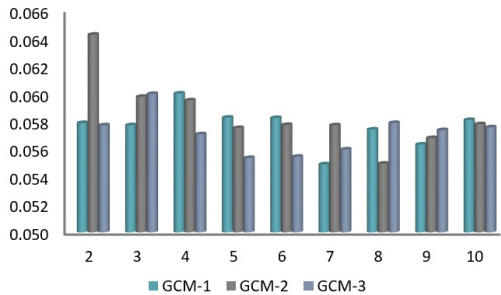


Fig. 11. NMSE as a function of the number of clusters c for the Ecoli dataset

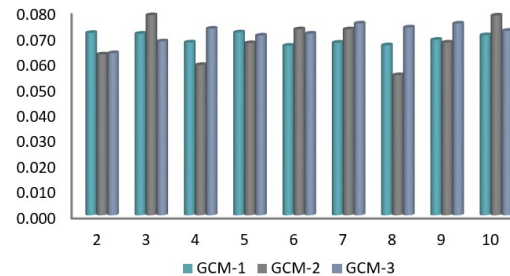


Fig. 15. NMSE as a function of the number of clusters c for the Appendicitis dataset

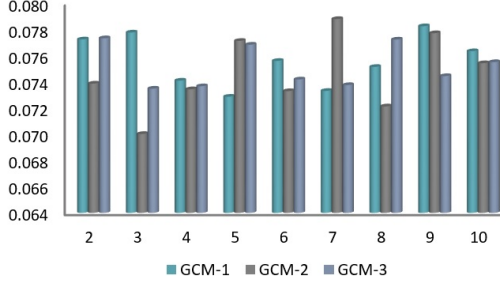


Fig. 16. NMSE as a function of the number of clusters c for the Glass dataset

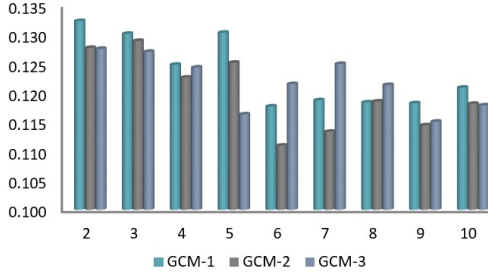


Fig. 17. NMSE as a function of the number of clusters c for the Glass0 dataset

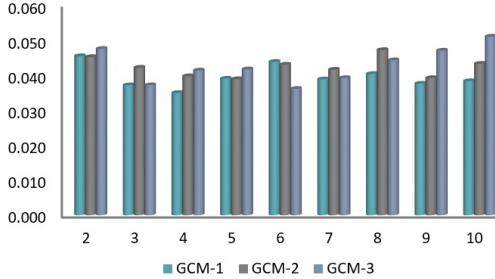


Fig. 18. NMSE as a function of the number of clusters c for the Glass2 dataset

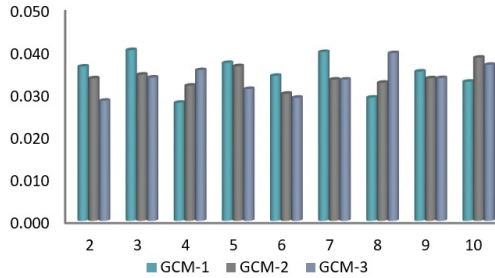


Fig. 19. NMSE as a function of the number of clusters c for the Glass3 dataset

The simulation results illustrate the strength of the proposed PGCM scheme to deal with GMLC instances. For example, the reader may observe that the NMSE value always falls below 0.1 for a specific number of clusters, barring the Glass0 altered dataset wherein the best performing model (GCM-2 with $c = 6$) reports an error slightly superior to 10%. This means that our three granular classifiers managed to efficiently compute the degree of each label from a set of numerical attributes. Given these numerical simulations, we can enunciate some concluding remarks:

- The performance among all neural models is soundly similar (the Friedman test [37] advocates the absence of significant differences among them at the 95% level, p -value = 0.59588). However, the second model achieved the best performance in 13/16 datasets. This suggests that including the inter-label correlation in the FCM topology may be potentially useful even when the method used for generating the synthetic datasets does not ensure the existence of such correlation.
- The performance of each granular model changes with the number of clusters. Increasing the granularity level does not inexorably imply higher accuracy. This fact leads us to believe that the ideal granularity level is actually problem-specific. Despite this realization, further measures for estimating this parameter could be adopted.

Another element to be discussed is the role of the clustering method that induces the granular FCMs. During empirical simulations we observed that the proposed model is occasionally sensitive to the initialization of the cluster prototypes. This is why the aforementioned results comprise the average of 10 independent trials. It suggests that more consistent initialization strategies must be adopted for boosting the PGCM performance. Moreover, the stochastic nature of the PSO algorithm may be responsible for the small variations on the NMSE measure. Nevertheless, the averaged outcomes illustrate the capability of the proposed granular systems for coping with GMLC problems described by a set of numerical features.

V. CONCLUSION

In this paper we have presented three implementations of a granular neural network model to solve GMLC scenarios. They inherit the capability of fuzzy clustering methods to discover information granules from data and the semantics of cognitive maps to infer new knowledge from concepts by using a neural reasoning rule. In this kind of granular system, fuzzy cluster prototypes are represented as sigmoid input-type neurons whereas output-type neurons denote problem labels. Furthermore, causal weights are automatically estimated from historical data using a supervised learning scheme powered by PSO as the underlying optimization engine. To activate the model we use the membership grade of the target object to each PGCM input neuron (cluster prototype).

Numerical experiments have confirmed the ability of these granular FCMs to accurately estimate the degree of association between an object and each label. The lack of

proper GMLC datasets in the literature led us to devise a simple procedure that transforms standard Machine Learning datasets into GMLC ones. However, the lack of pure GMLC algorithms makes it difficult to perform a sound statistical analysis between our method and those previously published. More explicitly, existing GMLC algorithms are focused on predicting the ordinal relation of labels instead of the exact membership grade to each label. We could adapt our proposal to the existing scenarios but the comparison then becomes unfair. The empirical analysis also suggests that all three granular models perform comparably; however this finding may not hold in the presence of a wider GMLC testbed. A more rigorous statistical analysis must be conducted to identify the strengths of each granular model topology.

Future research will be geared towards assessing the effect of the FCM convergence upon the predicted results, improving the initialization of the fuzzy cluster prototypes and developing PGCM extensions for handling datasets with nominal and mixed-type attributes.

ACKNOWLEDGMENT

We are grateful to Isel Grau, a PhD student at Vrije Universiteit Brussel (Belgium), for her valuable support and insightful comments during the design and implementation of this proposal.

REFERENCES

- [1] L. H. Ederington, "Classification models and bond ratings," *Financial Review*, vol. 20, no. 4, pp. 237–262, 1985.
- [2] S. B. Kotsiantis, I. Zaharakis, and P. Pintelas, "Supervised machine learning: a review of classification techniques," in *Emerging Artificial Intelligence Applications in Computer Engineering*. IOS Press, 2007, ch. 1, pp. 3–24.
- [3] G. Tsoumakas and I. Katakis, "Multi-label classification: An overview," *Dept. of Informatics, Aristotle University of Thessaloniki, Greece*, 2006.
- [4] G. Tsoumakas, I. Katakis, and I. Vlahavas, "Mining multi-label data," in *Data mining and knowledge discovery handbook*. Springer, 2010, pp. 667–685.
- [5] W. Cheng, E. Hüllermeier, and K. J. Dembczynski, "Graded multilabel classification: the ordinal case," in *Proceedings of the 27th international conference on machine learning (ICML-10)*, 2010, pp. 223–230.
- [6] R. Bello, R. Falcon, W. Pedrycz, and J. Kacprzyk, *Granular Computing: at the Junction of Rough Sets and Fuzzy Sets*. Berlin-Heidelberg, Germany: Springer Verlag, 2008.
- [7] A. Bargiela and W. Pedrycz, *Granular computing: an introduction*. Springer Science & Business Media, 2012, vol. 717.
- [8] W. Pedrycz, *Granular computing: Analysis and design of intelligent systems*. CRC Press, 2013.
- [9] L. Polkowski and P. Artienjew, "Granular computing: Granular classifiers and missing values," in *Cognitive Informatics, 6th IEEE International Conference on*. IEEE, 2007, pp. 186–194.
- [10] V. G. Kaburlasos and S. E. Papadakis, "A granular extension of the fuzzy-ARTMAP (FAM) neural classifier based on fuzzy lattice reasoning (FLR)," *Neurocomputing*, vol. 72, no. 10, pp. 2067–2078, 2009.
- [11] R. Al-Hmouz, W. Pedrycz, A. Balamash, and A. Morfeq, "From data to granular data and granular classifiers," in *Fuzzy Systems (FUZZ-IEEE), 2014 IEEE International Conference on*. IEEE, 2014, pp. 432–438.
- [12] G. Nápoles, I. Grau, E. Papageorgiou, R. Bello, and K. Vanhoof, "Rough cognitive networks," *Knowledge-Based Systems*, 2015.
- [13] G. Nápoles, I. Grau, R. Falcon, R. Bello, and K. Vanhoof, "A granular intrusion detection system using rough cognitive networks," in *Recent Advances in Computational Intelligence in Defense and Security*, R. Abielmona, R. Falcon, N. Zincir-Heywood, and H. Abbass, Eds. Springer Verlag, 2015, ch. 7.
- [14] M. Szczuka, A. Jankowski, A. Skowron, and D. Slezak, "Building granular systems - from concepts to applications," in *Rough Sets, Fuzzy Sets, Data Mining, and Granular Computing*, ser. Lecture Notes in Computer Science, Y. Yao, Q. Hu, H. Yu, and J. W. Grzymala-Busse, Eds. Springer International Publishing, 2015, vol. 9437, pp. 245–255.
- [15] N. Marz and J. Warren, *Big Data: Principles and best practices of scalable realtime data systems*. Manning Publications Co., 2015.
- [16] J. C. Bezdek, R. Ehrlich, and W. Full, "FCM: The fuzzy c-means clustering algorithm," *Computers & Geosciences*, vol. 10, no. 2, pp. 191–203, 1984.
- [17] C. Brinker, E. L. Mencia, and J. Furnkranz, "Graded multilabel classification by pairwise comparisons," in *Data Mining (ICDM), 2014 IEEE International Conference on*. IEEE, 2014, pp. 731–736.
- [18] G. Lastra, O. Luaces, and A. Bahamonde, "Interval prediction for graded multi-label classification," *Pattern Recognition Letters*, vol. 49, pp. 171–176, 2014.
- [19] B. Kosko, "Fuzzy cognitive maps," *International Journal of Man-Machine Studies*, vol. 24, no. 1, pp. 65–75, Jan. 1986.
- [20] —, *Fuzzy engineering*. Prentice-Hall, Inc., Oct. 1996.
- [21] S. Bueno and J. L. Salmeron, "Benchmarking main activation functions in fuzzy cognitive maps," *Expert Systems with Applications*, vol. 36, no. 3, pp. 5221–5229, Apr. 2009.
- [22] B. Kosko, "Hidden patterns in combined and adaptive knowledge networks," *International Journal of Approximate Reasoning*, vol. 2, no. 4, pp. 377–393, Oct. 1988.
- [23] W. Pedrycz and W. Homenda, "From fuzzy cognitive maps to granular cognitive maps," *Fuzzy Systems, IEEE Transactions on*, vol. 22, no. 4, pp. 859–869, 2014.
- [24] W. Pedrycz, "The design of cognitive maps: A study in synergy of granular computing and evolutionary optimization," *Expert Systems with Applications*, vol. 37, no. 10, pp. 7288–7294, 2010.
- [25] W. Pedrycz, A. Jastrzebska, and W. Homenda, "Design of fuzzy cognitive maps for modeling time series," *Fuzzy Systems, IEEE Transactions on*, vol. PP, no. 99, pp. 1–1, 2015.
- [26] W. Homenda, A. Jastrzebska, and W. Pedrycz, "Granular cognitive maps reconstruction," in *Fuzzy Systems (FUZZ-IEEE), 2014 IEEE International Conference on*. IEEE, 2014, pp. 2572–2579.
- [27] A. Abraham, R. Falcon, and R. Bello, *Rough Set Theory: a True Landmark in Data Analysis*. Berlin-Heidelberg, Germany: Springer Verlag, 2009.
- [28] E. Papageorgiou *et al.*, "Learning algorithms for fuzzy cognitive maps - a review study," *Systems, Man, and Cybernetics, Part C: Applications and Reviews, IEEE Transactions on*, vol. 42, no. 2, pp. 150–163, 2012.
- [29] G. Nápoles, R. Bello, and K. Vanhoof, "Learning stability features on sigmoid fuzzy cognitive maps through a swarm intelligence approach," in *Progress in Pattern Recognition, Image Analysis, Computer Vision, and Applications*. Springer, 2013, pp. 270–277.
- [30] —, "How to improve the convergence on sigmoid fuzzy cognitive maps?" *Intelligent Data Analysis*, vol. 18, no. 6S, pp. S77–S88, 2014.
- [31] —, "On the convergence of sigmoid Fuzzy Cognitive Maps," *Information Sciences*, submitted.
- [32] R. Poli, J. Kennedy, and T. Blackwell, "Particle swarm optimization - a survey," *Swarm intelligence*, vol. 1, no. 1, pp. 33–57, 2007.
- [33] G. Nápoles, I. Grau, M. Bello, and R. Bello, "Towards swarm diversity: Random sampling in variable neighborhoods procedure using a lévy distribution," *Computación y Sistemas*, vol. 18, no. 1, pp. 79–95, 2014.
- [34] A. E. Abele-Brehm and M. Stief, "Die prognose des berufserfolgs von hochschulabsolventinnen und-absolventen: Befunde zur ersten und zweiten erhebung der erlanger längsschnittstudie bela-e," *Zeitschrift für Arbeits-und Organisationspsychologie A&O*, vol. 48, no. 1, pp. 4–16, 2004.
- [35] L. Breiman, "Random forests," *Machine learning*, vol. 45, no. 1, pp. 5–32, 2001.
- [36] M. Clerc and J. Kennedy, "The particle swarm-explosion, stability, and convergence in a multidimensional complex space," *Evolutionary Computation, IEEE Transactions on*, vol. 6, no. 1, pp. 58–73, 2002.
- [37] M. Friedman, "The use of ranks to avoid the assumption of normality implicit in the analysis of variance," *Journal of the American Statistical Association*, vol. 32, no. 200, pp. 675–701, 1937.