Neural Networks 97 (2018) 19–27

Contents lists available at ScienceDirect

Neural Networks

journal homepage: www.elsevier.com/locate/neunet

Fuzzy-Rough Cognitive Networks

Gonzalo Nápoles [a,*], Carlos Mosquera [b], Rafael Falcon [c], Isel Grau [b], Rafael Bello [b], Koen Vanhoof [a]

[a] Faculty of Business Economics, Universiteit Hasselt, Belgium
[b] Department of Computer Sciences, Central University of Las Villas, Cuba
[c] School of Electrical Engineering and Computer Science, University of Ottawa, Canada

ARTICLE INFO

Article history:
Received 27 January 2017
Received in revised form 20 August 2017
Accepted 22 August 2017
Available online 28 September 2017

Keywords:
Fuzzy cognitive maps
Fuzzy rough sets
Rough cognitive mapping
Pattern classification
Granular classifiers

ABSTRACT

Rough Cognitive Networks (RCNs) are a kind of granular neural network that augments the reasoning rule present in Fuzzy Cognitive Maps with crisp information granules coming from Rough Set Theory. While RCNs have shown promise in solving different classification problems, this model is still very sensitive to the similarity threshold upon which the rough information granules are built. In this paper, we cast the RCN model within the framework of *fuzzy rough sets* in an attempt to eliminate the need for a user-specified similarity threshold while retaining the model's discriminatory power. As far as we know, this is the first study that brings fuzzy sets into the domain of rough cognitive mapping. Numerical results in the presence of 140 well-known pattern classification problems reveal that our approach, referred to as *Fuzzy-Rough Cognitive Networks*, is capable of outperforming most traditional classifiers used for benchmarking purposes. Furthermore, we explore the impact of using different heterogeneous distance functions and fuzzy operators over the performance of our granular neural network.

1. Introduction

Pattern classification (Duda, Hart, & Stork, 2012) is one of the most popular field within Artificial Intelligence as a result of its link with real-world problems. In short, it may be defined as the process of identifying the right category (among those in a predefined set) to which an observation belongs. The ease with which we recognize our beloved black cat from hundreds similar to it or read handwritten characters belies the astoundingly complex processes that underlie these scenarios. That is why researchers have been focused on developing a wide spectrum of classification algorithms called *classifiers* with the goal of solving these problems with the best possible accuracy.

The literature on classification models (Witten & Frank, 2005) is vast and offers a myriad of techniques that approach the classification problem from multiple angles. Regrettably, some of the most accurate classifiers do not provide any mechanism to explain how they arrived at each conclusion and behave like *black boxes*. This means that their reasoning mechanism is not transparent, therefore negatively affecting their practical usability in scenarios where understanding the decision process is required. According to the terminology discussed in Nápoles (2017), *transparency* can be understood as the classifier's ability to explain its reasoning mechanism, whereas *interpretability* refers to the classifier's ability to explain the problem domain at the attribute level.

Recently, Nápoles and his collaborators (Nápoles, Grau, Papageorgiou, Bello, & Vanhoof, 2016) introduced the *Rough Cognitive Networks* (RCNs) in an attempt to develop an accurate, transparent classifier. Such granular neural networks augment the reasoning scheme present in Fuzzy Cognitive Maps (FCMs) (Kosko, 1986) with information granules coming from Rough Set Theory (RST) (Pawlak, 1982). Although RCNs can be considered as recurrent neural systems that fit the McCulloch–Pitts' scheme (McCulloch & Pitts, 1988), there are important differences with regards to other neural models.

Classical neural networks regularly perform like black boxes, where neither neurons nor connections have any clear specific meaning for the problem itself (Nápoles, Papageorgiou, Bello, & Vanhoof, 2016). However, all the neurons and connections in an RCN have a precise meaning at a granular level, therefore making it possible to understand the underlying decision process at a granular (symbolic) level. The absence of hidden neurons and the lazy learning approach are other distinctive features attached to these granular, recurrent neural systems.

While RCNs have shown promise in solving different pattern classification problems (Nápoles, Grau, Falcon, Bello, & Vanhoof, 2016; Nápoles,Grau, Papageorgiou et al., 2016), their performance

* Corresponding author.
E-mail addresses: gonzalo.napoles@uhasselt.be (G. Nápoles), xcarlos@gmail.com (C. Mosquera), rfalc032@uottawa.ca (R. Falcon), igrau@uclv.edu.cu (I. Grau), rbellop@uclv.edu.cu (R. Bello), koen.vanhoof@uhasselt.be (K. Vanhoof).

https://doi.org/10.1016/j.neunet.2017.08.007
0893-6080/

is still very sensitive to an input parameter denoting the similarity threshold upon which the rough information granules are built. The proper estimation of this parameter is essential in the presence of numerical attributes since it defines whether two objects are deemed similar or not. Aiming at overcoming this drawback, Nápoles,Grau, Papageorgiou et al. (2016) proposed an optimization-based hyperparameter learning scheme to estimate the value of this parameter from a hold-out test set. However, this strategy may become impractical for large datasets since it requires rebuilding the information granules for each parameter value to be evaluated.

In Nápoles, Falcon, Papageorgiou, Bello, and Vanhoof (2017) the authors proposed a granular ensemble named *Rough Cognitive Ensembles* (RCEs) to deal with the parametric requirements of RCN-based classifiers. This classification model employs a collection of RCNs, each operating at a different granularity degree. While this approach involves a more elaborated solution, the ensemble architecture and the bagging strategy used to improved the diversity among the base classifiers irremediably harm the transparency of RCNs, thus becoming another black-box.

In this paper, we cast the RCN approach within the framework of Fuzzy Rough Set Theory (FRST) (Cornelis, De Cock, & Radzikowska, 2008; Dubois & Prade, 1990; Inuiguchi, Wu, Cornelis, & Verbiest, 2015; Radzikowska & Kerre, 2002) in an attempt to eliminate the need for a user-specified similarity threshold while retaining the model's discriminatory power. Fuzzy rough sets are an extension of classical rough sets in which fuzzy sets are used to characterize the degree to which an object belongs to each information granule. The inclusion of the fuzzy approach into the RCN model allows coping with both the vagueness (fuzzy sets) and inconsistency (rough sets) of the information typically found in pattern classification environments. Besides, it allows designing a more elegant solution for the parametric issues of RCN-based classifiers.

Numerical simulations using 140 datasets reveal that the proposed model, referred to as *Fuzzy-Rough Cognitive Networks* (FRCNs), is capable of outperforming the standard RCNs using a fixed, reasonable similarity threshold value. The results also suggest that FRCNs remain competitive with regards to RCEs and other black-box classifiers adopted for comparison purposes. More importantly, the challenging process of estimating a precise value for the similarity threshold parameter is no longer a concern.

The remainder of this paper is organized as follows. Section 2 briefly describes the RCN algorithm and the motivation behind our proposal. The fuzzy RCN classifier is unveiled in Section 3, whereas Section 4 introduces the numerical simulations and their ensuing discussion. Towards the end, Section 5 outlines some concluding remarks and future work directions.

2. Rough cognitive mapping

This section discusses the technical background relevant to this study and explains the motivation behind the fuzzy approach.

2.1. Theoretical background

Rough cognitive mapping is a recently introduced concept (Nápoles,Grau, Papageorgiou et al., 2016) that brings together RST and FCMs. RCNs are granular FCMs whose topology is defined by the abstract semantics of the three-way decision rules (Yao, 2009, 2011). The set of input neurons in an RCN represent the positive, boundary and negative regions of the decision classes in the problem under consideration. The output neurons describe the set of decision classes. The topology (both concepts and weights) is entirely computed from historical data, thus removing the need for expert intervention during the classifier's construction.

The first step in the RCN learning process is related to the *input data granulation* using RST. The positive, boundary and negative regions of each decision class according to a subset of attributes are computed using the training data set and a predefined similarity relation.

The second step is concerned with *topology design* where a sigmoid FCM is automatically created from the discovered information granules by using a set of predefined rules; see Nápoles,Grau, Papageorgiou et al. (2016) for more details. In principle, an RCN will be composed of at most $4|\mathcal{D}|$ neurons and $3|\mathcal{D}|(1+|D|)$ causal relationships, with $\mathcal{D} = \{D_1, \dots, D_K\}$ being the set of decision classes.

The last step refers to the *network exploitation*, which simply means computing the response vector $\mathcal{A}_x(\mathcal{D}) = \{A_x(D_1), \dots, A_x(D_k), \dots, A_x(D_K)\}$ for some unlabeled object. The new object x is presented to the RCN as an input vector $A^{(0)}$ that activates input neurons. Each element in $A^{(0)}$ is computed on the basis of the inclusion degree of x to each rough granular region. After this, the input vector is propagated through the RCN using the McCulloch–Pitts reasoning model (McCulloch & Pitts, 1988) and next the decision class with the highest value in the response vector is then assigned to the test object.

2.2. Motivation for the FRCN approach

The notion of *rough cognitive mapping* opened up a new research avenue in the field of granular-neural classifiers. However, their performance is highly sensitive to the similarity threshold used to determine whether two instances can be gathered together into the same similarity class.

Nápoles,Grau, Papageorgiou et al. (2016) used a parameter tuning method based on the Harmony Search (HS) optimizer to estimate the similarity threshold. Nevertheless, the evaluation of every candidate solution requires recalculating the lower and upper approximations of each RST-based region for each decision class, which could be computationally prohibitive for large datasets.

Let us assume that $\mathcal{U}_1 \subset \mathcal{U}$ is the training set and $\mathcal{U}_2 \subset \mathcal{U}$ is the hold-out test (validation) set such that $\mathcal{U}_1 \cap \mathcal{U}_2 = \emptyset$. The computational complexity of building the lower and upper approximations is $O(|\Phi|\,|\mathcal{U}_1|^2)$, with Φ being the attribute set, whereas the complexity of building the network topology is $O(|\mathcal{D}|^2)$, with $\mathcal{D}$ being the set of decision classes. Besides, the complexity of exploiting the granular network for $|\mathcal{U}_2|$ instances is $O(|\mathcal{U}_2\|\Phi\|\mathcal{U}_1|^2)$. This implies that the temporal complexity of evaluating a single parameter value is $O(\max\{|\Phi|\,|\mathcal{U}_1|^2, |\mathcal{D}|^2, |\mathcal{U}_2\|\Phi\|\mathcal{U}_1|^2\})$. Due to the fact that $|\mathcal{U}_1| \geq |\mathcal{U}_2|$ in most machine learning scenarios, we can conclude that the overall complexity of this parameter learning method is $O(T|\Phi|\,|\mathcal{U}_1|^3)$, where T is the number of learning cycles. Regrettably, this may negatively affect the practical usability of RCNs in solving the real-world pattern classification problems.

The key goal behind this research is to remove the estimation of the similarity threshold without affecting the overall RCN's discriminatory power. Being more explicit, we aim to arrive at a parameterless classifier (and hence suppressing the need for a parameter tuning strategy) without degrading the RCN's performance in classification problems.

3. Fuzzy-rough cognitive mapping

This section presents the notion of *fuzzy-rough cognitive mapping* in order to remove the requirement of estimating the similarity threshold in an RCN. With this goal in mind, we first describe the mathematical foundations behind this approach. Afterwards, we explain how to construct an FRCN for solving pattern classification problems.