DIMENSIONALITY REDUCTION USING ROUGH SET APPROACH FOR CLIMATE PREDICTION

Juliana Aparecida Anochi *
*National Institute for Space Research (INPE)**
*São José dos Campos, SP, Brazil **

Haroldo Fraga de Campos Velho *
*National Institute for Space Research (INPE)**
*São José dos Campos, SP, Brazil **

ABSTRACT

In this article, a data mining method to variables selection for climate prediction is presented. The data were processed by Rough Set Theory to extract relevant information to perform the seasonal climate prediction by neural network for the South of Brazil, with a reduced data set. The neural network was self-configured by MPCA metaheuristic. Two experiments were conducted with neural network: complete meteorological input variables, and reduced data set extract from the rough set theory.

KEYWORDS

Data mining, rough sets, optimal neural network, multiple particle collision algorithm, climate prediction.

1. INTRODUCTION

A technique for data reductions is presented. The determination of meteorological variables is processed by a Rough Sets Theory (RST) to generate data reductions, in order to adjust the neural network models for prediction seasonal precipitation, given the hypothesis that some variables are more relevant to climatic forecasting process, and a data mining technique can identify these variables.

Considering the possibility to develop climate prediction models from data describing the behavior of weather conditions. In this context, the method must need a large number of data to ensure that the model considers a wide range of situations. However, despite ensure at first robustness to derivative models, handling of a large amounts of data may require a high computational complexity, preventing the use of the development methodology for calibrating local models.

Rough set theory is one of many methods that can be employed to treat uncertain and imprecise information, by deriving approximations of a data set. The selected attributes are used as inputs to the neural network (NN) learning process (Pawlak, 1982). Many researches have used the RST for selecting attributes to use as inputs to the neural networks. Chun-Yan et al. (2003) proposes a method to combine rough set theory with neural network for pattern recognition, and compares the results with the standard methods. Jiang et al. (2008) proposes a method for image segmentation based on RST and NN: the rough set is used to reduce the image attribute.

Despite the potential of a neural network model, its performance is dependent on the topology (architecture), with significant influence in the training process. The definition of the topology of a supervised neural network depends on the following parameters: the number of hidden layers (intermediate), the number of neurons in each layer, type of activation function, learning rate and momentum rate.

Increased attention has been especially directed to finding the best architecture (Benardos et al., 2007; Carvalho et al., 2012; Kordik et al., 2010). Doulamis et al. (2003) have proposed adaptable neural network where the weights are up-dated on demand. Here, the goal is to identify the best configuration for the applied neural network. This is a complex task, and usually requires a great effort by an expert, determining the best parameter set, and it is necessary a previous knowledge about the problem to be treated. In this article, the determination of optimal parameters is formulated as an optimization problem, solved by the metaheuristic Multiple Particle Collision Algorithm (MPCA). The application is focused on developing an empirical model of seasonal climate prediction precipitation for the Brazilian regions. The technique will be illustrated for the Brazilian South region.

2. ROUGH SET THEORY

Rough Sets Theory (RST) was proposed in 1982 by Zdzislaw Pawlak as a mathematical theory to treat uncertain and imprecise information, by deriving approximations of a data set (Pawlak, 1982). The fundamental feature behind rough set is the approximation of lower and upper spaces of a set, the approximation of spaces being the formal classification of knowledge regarding the interest domain.

Rough Sets Theory uses the concept of Information Systems (IS) in which the available data are represented in a table in which the objects are displayed in the rows and the attributes in the columns (Komorowski and Ohrn, 1999). Formally, an information system is composed of a finite non-empty set U (Universe) of objects and a finite non-empty set A of attributes, $SD=(U,A)$, such that $a:U \rightarrow V_a$, for every $a \in A$. The set V_a is called the value set of a. A decision system (DS) is any information system of the form $DS=(U, A \cup \{d\})$, where $d \notin A$ is the *decision attribute.*

The indiscernibility relation is used as a measure of similarity among objects. Thus, a set of objects with the same attributes are indiscernible if only if their attributes hold the same values from their corresponding domains. This equivalence relation may be used to treat problems as redundancy of attributes or the existence of irrelevant attributes in the data assigned to only one representative of a class.

The reduction process is identify equivalence classes, i.e. objects that are indiscernible using the available attributes. Savings are to made since only one element of the equivalence class is needed to represent the entire class. The other dimension in reduction is to keep only those attributes that preserve the indiscernibility relation and, consequently, set approximation (Komorowski and Ohrn, 1999). Thangavel and Pethalakshmi (2009) did a review of different methodologies to implement the RST strategy. Here, the Rosetta package (http://www.lcb.uu.se/tools/rosetta) will be used.

3. METHODOLOGY

Artificial Neural Networks (ANN) are computational techniques that present a mathematical model inspired by the neural structure of biological organisms, acquiring knowledge through experience. The intelligent behavior emerges from the interactions between processing units, from their environment through a learning process, whose function is to modify the synaptic weights of the network and provides this knowledge to the application in question. The recurrent Elman is the neural networks used here.

The Elman network (Elman, 1990) contains recurrent connections from the hidden neurons to a layer of context units consisting of unit delays. The context unit store the outputs of the hidden neurons for one time step, and then feed them back to the input layer. The hidden neurons thus have some record of their prior activation, which enables the network to perform learning tasks that extend over time.

The parameter setting problem of an optimal architecture for a neural network is described as an optimization problem where each point in the search space represents a different topology. The definition of Elman recurrent network topology, which is the object of this work basically deals the following parameters: number of hidden layers (intermediate), the number of neurons in each layer, type of activation function, learning rate and momentum constant. The objective function used in this study is the square difference between the target values and the NN output. The cost function is given by (Carvalho et al., 2011).

$$f_{obj} = penalty \times \left(\frac{\rho 1 \times E_{train} + \rho 2 \times E_{gen}}{\rho 1 + \rho 2} \right) \tag{5}$$

where $\rho 1 = 1$ and $\rho 2 = 0.1$, the same values proposed by (Carvalho et al., 2011). These are adjustment factors for taking into account the relative relevance of training and generalization errors.

The factor *penalty* is applied to compute the ANN with the lowest complexity possible. Here, the computational complexity of Elman-NN architecture can be defined as the total number of neurons and the epochs number present in its structure. Thus, a penalty factor was developed to favour lightweight architectures, and it can be expressed by:

$$penalty = C_1\left(\varepsilon^{neurons}\right)^2 + C_2\left(epochs\right) + 1 \tag{6}$$

where C_1=1 and C_2=0.1 are adjustment parameters for finding a good balance for the two factors in the complexity measurement.

The RST was used to extract relevant information from meteorological data to perform climate prediction, from a reduced data set. In the dimensionality reduction process the relevant attributes are those that mostly occur in the data, in terms of the indiscernibility relation. For the attribute reduction process the training data set previously mentioned was first discretized and then submitted to the reduction algorithm for selection of the relevant attributes chosen as those with a presence greater than 70% in the discernibility function.

3.1 Multiple Particle Collision Algorithm

The Multiple Particle Collision Algorithm (MPCA) was developed by Luz et al. (2008), inspired in the canonical Particle Collision Algorithm (PCA) (Sacco et al., 2006). In this version of the algorithm, a new characteristic is introduced: the use of several particles, instead of only one particle, for moving in the search space. The metaheuristic PCA was inspired on the phenomena occurring inside of a nuclear reactor during a neutron travel, specifically in two physical behaviors, absorption (particle is captured by the target core) and scattering (the particle follows different direction after collision with the target).

The PCA starts with a selection of an initial solution (Old-Config). It is modified by a stochastic perturbation (function Perturbation), leading to the construction of a new solution (New-Config). The new solution is compared (function Fitness), and the new solution can or cannot be accepted. If the new solution is not accepted, a Metropolis' scheme (Metropolis et al., 1953) is used: the particle can be send to a different place in the search space, giving the algorithm the capability of escaping a local optimum. This approach is inspired on the scattering process (function Scattering). If a new solution is better than the previous one, this new solution is absorbed (function Absorption).

4. EXPERIMENTS AND RESULTS

Climate prediction is the estimation of the average behavior of the atmosphere for a future period of time (more than one month ahead). Two models of Elman recurrent networks were constructed: one model was generated using all available variables in the database, with the optimization parameters by using the metaheuristic MPCA. The other model, containing a reduced number of variables indicated by RST, using self-configuring of parameters by MPCA.

The selected study area was South Brazil (Lat 35°S, 25°S) to (Lon 60°W, 40°W). The data consists of monthly means from January 1990 to February 2015. The data was downloaded from the reanalysis data repository from National Center for Environmental Prediction & National Center for Atmospheric Research (NCEP/NCAR). The global data reanalysis grid uses a set with horizontal resolution of 2.5° x 2.5° (latitude x longitude). The available variables are presented: temperature (temp), zonal wind components at levels 300, 500, 850 hPa, meridional wind components at levels 300, 500, 850 hPa, specific humidity (shum) and precipitation (prec).

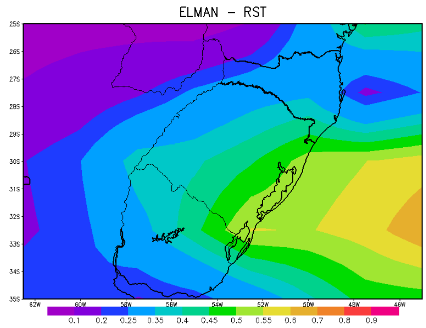
The training data subset was formed with data from January 1990 up to December 2011. This subset was used to derive a NN model. The period from January 2012 to December 2012 was used for the validation. The generalization test subset corresponds to the period from January 2013 up to January 2015. The variables that form the products has a presence greater than 70% in the discernibility function. It can be noticed that only 6 variables out of 9, were considered. The variables selected by RST are: temp (77%), u300 (79%), u850 (80%), v300 (77%), v500 (79%) and v850 (70%).

Table 1. Configuration topology the Elman-NN

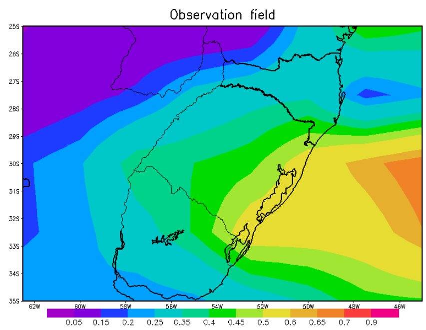
Parameters	Elman-NN
Intermediate layer	1
Neuron in first layer	15
Learning rate	0.53
Momentum rate	0.24
Activation function	Logistics

4.1 Results for climate precipitation prediction

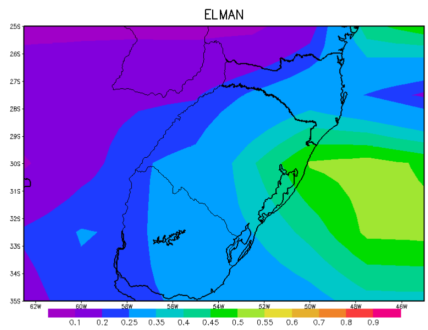
Figure 1 shows the results obtained for climate prediction in South of Brazil using the Elman-NN for the season winter. Figure 1a corresponds to the observed field in winter of 2014; Figure 1b shows the result produced by the Elman-NN trained with a reduced set, and the topology was self-configured by MPCA; Figure 1c shows the result obtained by the Elman-NN that was design by MPCA and was trained with all available variables in the database. The differences between observation and forecasted precipitation are presented in Figures 1d and 1e, with different NNs, trained with a reduced set and trained with all available variables, respectively.



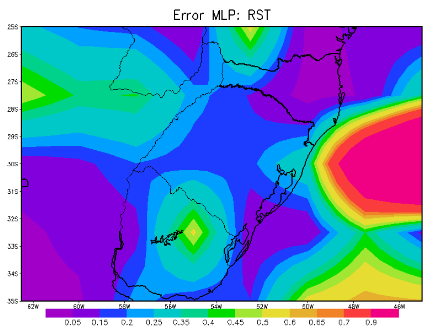
(a) Observation during the winter-2014



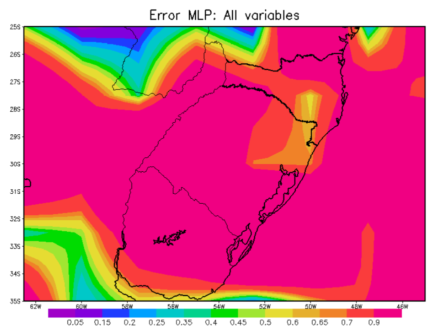
(b) Elman-RST: MPCA



(c) Elman-complete variables



(d) Error-RST: MPCA



(e) Error-variables complete

Figure 1: Climate prediction for winter in 2014

From the results, the Elman network trained with the set reduced by RST present the better performance than the models trained with all available variables (see Figure 1d).

5. CONCLUSION

This work presented an approach to determine optimal neural network architecture, with input variables selected by RST to seasonal climate precipitation prediction. The Elman recurrent neural network was employed. The RST approach was able to reduce the observation data dimension. The use of reduce input attributes produced a better prediction than a full meteorological variables available. The self-configured with reduce input provided good results for climate precipitation prediction for South Brazilian region.

The advantage of using an automatic procedure to configure an ANN is no needing of helping from an expert on the NN approach and/or on the application, in order to have a functional neural network.

ACKNOWLEDGEMENT

Authors would like to thank CNPq and CAPES, Brazilian agencies for research support.

REFERENCES

Benardos, P. and Vosniakos, G.-C, 2007. Optimizing feedforward artificial neural network architecture. *Engineering Applications of Artificial Intelligence,* Vol. 20, No. 3, pp. 365-382.

Carvalho, A., Ramos, F.M., and Chaves, A.C., 2011. Metaheuristics for the feedforward artificial neural network (ANN) architecture optimization problem. *Neural Comput Applic*, Vol. 20, pp. 1273–1284.

Chun-Yan, Y., Ming-Hui, W. and Ming, W., 2003. Combining rough set theory with neural network theory for pattern recognition. *Proceedings in IEEE International Conference on Robotics, Intelligent Systems and Signal Processing*. Vol. 2, pp. 880-885.

Doulamis, A.D., Doulamis, N.D., and Kollias, S.D., 2003. An adaptable neural-network model for recursive nonlinear traffic prediction and modeling of MPEG video sources. *Neural Networks*, IEEE Transactions on, Vol. 14, No. 1, pp. 150-166.

Elman, J.L., 1990. Finding structure in time. *Cognitive Science*, Vo. 14, pp. 179–211.

Jiang, J., Yang, D., and Wei, H., 2008. Image segmentation based on rough set theory and neural networks. *Proceedings of International Conference on Visual Information Engineering*. China. pp. 361-365.

Kordík, P., Koutník, J., Drchal, J., Kovářík, O., Čepek, M. and Šnorek, M., 2010. Meta-learning approach to neural network optimization. *Neural Networks*. Vol. 23 No. 4, pp. 568-582.

Komorowski, J. and A Ohrn, 1999. Modelling prognostic power of cardiac tests using rough sets, *Artificial Intelligence in Medicine*, Vol. 15, pp. 167-191.

Luz, E.F.P., Becceneri, J.C., and Campos Velho, H.F., 2008. A new multi-particle collision algorithm for optimization in a high-performance environment. *Journal of Computational Interdisciplinary Sciences*, Vol. 1, No. 1, pp. 3-10.

Metropolis, N., Resenbluth A.W., Rosenbluth, M.N., Teller, A.H. and Teller, E. 1953. Equation of state calculations by fast computing machines. *The Journal of Chemical Physics*. Vol. 6, No. 21, pp. 1087-1092.

Pawlak Z., 1982. Rough sets, *International Journal of Computer and Information Sciences*, Vol. 1, pp. 341-356.

Sacco, M., Lapa, C., Pereira, C., and Filho, H., 2006. Two stochastic optimization algorithms applied to nuclear reactor core design. *Progress in Nuclear Energy*, Vol. 48, No. 6, pp. 525–539.

Thangavel, K., & Pethalakshmi, A., 2009. Dimensionality reduction based on rough set theory: A review. Applied Soft Computing, 9(1), 1-12.