Climate precipitation prediction using optimal neural network architecture in Southeast Region of Brazil

Juliana A. Anochi[a 1], Haroldo F. de Campos Velho [b] and Reynier Hernández Torres[b]

[a]Center for Weather Forecasts and Climate Studies, Cachoeira Paulista, SP, Brazil
[b]National Institute for Space Research, São José dos Campos, SP, Brazil

Abstract

Neural network is a technique successfully employed in many applications on several research fields. Despite the potential of a neural network model, its performance is dependent on the definition of the parameters, since the definition of architecture (topology) can significantly influence the training process. Here, a technique for automatic configuration for a neural network is described as an optimization problem combining two different optimization schemes: a mono-objective minimization problem using Multi-Particle Collision Algorithm (MPCA), and a multiobjective minimization problem Nondominated Sorting Genetic Algorithm (NSGA-II). The proposed optimization approaches were tested for the mesoscale seasonal climate prediction for precipitation. The meteorological data were processed by Rough Set Theory to extract relevant information to perform the climate prediction by neural network for the Southeast region of Brazil, with a reduced data set.

Keywords: metaheuristic, optimization problem, neural networks, climate prediction, mono-objective problem, multiobjective problem.

1. Introduction

Climate precipitation prediction field is a key aspect in meteorology. The precipitation is a variable associated with agricultural crops, natural disasters (droughts and floods) with impacts on human begins, sectors of tourism and shipping. However, precipitation is difficult to predict because of large spatial and temporal variability (i.e. variable discontinuous). A method based on Artificial Neural Network (ANN) is applied to climate prediction precipitation in the Southeast region of Brazil. It is known the learning and response abilities of a neural network, which motivates their successful

[1]E-mail Corresponding Author: juliana.anochi@inpe.br

application in a wide variety of problems, consolidating its position as a solution technique for complex problems in pattern recognition, classification, control systems, proximity functions, and predictive model.

Despite the potential of a neural network model, its performance is dependent on the definition of the parameters, since the definition of architecture (topology) can significantly influence the training process. An appropriate configuration of a neural network is a complex task, and it often requires the knowledge of an expert on the application. In this context, a technique for automatic configuration for a neural network is described as an optimization problem combining two different optimization schemes: a mono-objective minimization problem using Multi-Particle Collision Algorithm (MPCA) [7], and a multiobjective minimization problem Non-dominated Sorting Genetic Algorithm (NSGA-II) [4].

In most scientific areas, there are many data sources. In meteorology, data from satellites, ground-based stations, ocean buoys, soundings, radar, and many others, are used in weather and climate forecasts. Predicting meteorological events is a complex challenge. The size reduction of the observation data without losing information is an important issue of research. For this reason, the Rough Sets Theory, a data mining technique, was used to identify the most significant variables for the climate prediction process [1].

The paper is organized as follows: Section 2 addresses rough sets for data mining; Section 3 deals with brief review on neural networks; Section 4 describes the study case adopted in this work; Section 5, the results and discussions are expressed; some final remarks are presented in Section 6.

2. Rough Sets in Data Analysis

The Rough Sets Theory (TCA) was proposed by Pawlak [10], as a mathematical approach for the treatment of uncertain and imperfect knowledge.

There are several formal models available for the treatment of uncertainties contained in a database, such as Fuzzy Sets Theory [13]; Dempster-Shafer theory (DST) [5] and Possibility Theory [14].

2.1 - Information and Decision System

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 [6].

Formally, an IS is composed of a finite non-empty set U (Universe) of objects and a finite non-empty set A of attributes, $IS = (U, A)$, so that, for each $a \in A$, $a : U \to V_a$. The set V_a is the set of values of a, that is, the domain of a. A Decision System (SD) is an IS augmented with a decision attribute $d \notin A$. Formally, $DS = (U, A\{d\})$, where $d \notin A$ is the decision attribute [6].

2.2 - Indiscernibility relation

The indiscernibility relation is used as a measure of similarity among objects. If two objects belong to the same equivalence class, they are indiscernible from each other. This is a equivalence relation that may be used to treat problems as redundancy of attributes [6].

$$I_A(B) = \left\{ (x, x') \in U^2 | \forall a \in B, a\,(x) = a\,(x') \right\}, \tag{1}$$

where $I_A(B)$ is the indiscernibility relation B. If $(x, x') \in I_A(B)$, then the objects x e x' are indiscernible for any attribute in the set B [6].

2.3 - Attribute Reduction

The reduction process is identify equivalence classes, i.e. objects that are indiscernible using the available attributes. The other dimension in reduction is to keep only those attributes that preserve the indiscernibility relation [6]. Here, the Rosetta package (http://www.lcb.uu.se/tools/rosetta) will be used.

3. Self-configured Neural Network

The use of meta-heuristics has been proposed as an efficient alternative to the configuration topologies parameters of NNs ([2], [3]). The definition of the architecture of a neural network can be formulated as an optimization problem, where each point in the search space represents an architecture.

In this paper, two formulations for automatic determination of optimal parameters of multilayer percetron network (NN) are employed for solving optimization problems: mono-objective minimization problem using Multi-Particle Collision Algorithm (MPCA), and a multiobjective minimization problem applying Non-dominated Sorting Genetic Algorithm (NSGA-II). In both problems, four mixed-variable domain parameters will be optimized: two continuous variables (the learning rate parameter η, and the momentum constant α) and two discrete variables (the number of neurons in the hidden layer (n_{neurons}), and the type of activation function ($f_{\text{activation}}$)).

The allowed values for these parameters are shown in Table 1.

3.1 - Mono-objective optimization by MPCA

Multi-Particle Collision Algorithm (MPCA) is a Metropolis type algorithm developed by Luz [7]. It is based on the Particle Collision Algorithm (PCA), introduced by Sacco [12]. The PCA is inspired by the physics of

Table 1: Parameters to define a network architecture

Type	Parameter	Value
Discrete	n_{neurons}	$[1, 40]$
	$f_{\text{activation}}$	tangent (1) sigmoid (2) gaussian (3)
Continuous	η	$[0.01, 1.0]$
	α	$[0.01, 1.0]$

travelling nuclear particle inside of a nuclear reactor, emulating the absorption (particle is captured by the target core) and scattering (the particle follows different direction after collision with the target).

The MPCA starts with a selection of an initial solution (`Old-Config`), and it is modified by a stochastic perturbation (`Perturbation`{.}), leading to the construction of a new solution (`New-Config`). The new solution is compared (by the function `Fitness`{.}), and the new solution can or cannot be accepted. If the new solution is not accepted, a Metropolis [9] scheme is used (the function `Scattering`{.}). The exploration on closer positions is guaranteed by using functions `Perturbation`{.}. If a new solution is better than the previous one, this new solution is absorbed `Aborption`{.} [7, 12]. The pseudocode for the MPCA is presented by Figure 1 – adapted from [7].

```
Generate an initial solution: Old-Config
Best-Fitness = Fitness{Old-Config}
Update Blackboard
For n = 0 to # of particles
    For n = 0 to # iterations
    Update Blackboard
    Perturbation{.}
        If Fitness{New-Config} > Fitness{Old-Config}
            If Fitness{New-Config} > Best-Fitness
                Best-Fitness = Fitness{New-Config}
            End If
            Old-Config = New-Config
            Exploration{.}
        Else
            Scattering{.}
        End If
    End For
End For
```

Figure 1: MPCA Pseudocode.

The objective function J of the mono-objective optimization problem, is the combination of two error criteria multiplied by a penalty factor (complexity of the ANN) architecture. The minimum value of J corresponds to the simple architecture and a good performance. This function is defined by [2].

$$J = penalty \times \frac{\rho_1 E_{train} + \rho_2 E_{\text{gen}}}{\rho_1 + \rho_2}, \tag{2}$$

where $\rho_1 = 1$ e $\rho_2 = 0.1$, the same values proposed in [2]. These are adjustment factors that modify the importance given to training and generalization errors, respectively.

The training error criterion (E_{train}) provides a quantitative indication of the neural network training level. The generalization error criterion (E_{gen}) refers to the neural network capacity to identify and respond similar patterns based on knowledge learned. The mean square error (MSE) is the metric used in these two criteria (E_{train} and E_{gen}), described in Equation 3.

$$MSE = \frac{1}{N} \sum_{k=1}^{N} (d_k - \hat{y}_k)^2 , \tag{3}$$

where N It is the number of training patterns, d_k is the desired output, and $\hat{y}_k$ the output is produced by the neural network.

The term *penalty* is a measure of computational complexity of the NN architecture, using the number of neurons present in the intermediate layer (n_{neurons}), and the total number of epochs (n_{epochs}), described in Equation 4.

$$penalty = C_1 e^{(n_{\text{neuron}})^2} + C_2(n_{\text{epochs}}) + 1, \tag{4}$$

where $C_1 = 5 \times 10^{-8}$ and $C_2 = 5 \times 10^{-5}$, are setting parameters to find an equilibrium in the measuring factors of complexity.

3.2 - Multi-objective optimization by NSGA-II

For multiobjective optimization, the goal is to calculate simultaneously optimal values for more than one objective function, more precisely, the idea is to identify non-dominated solutions. The set of non-dominated solutions defines the Pareto set. A desired response, the optimized topology, is obtained selecting one element from the Pareto set.

The mono-objective problem presented in previous subsection 3.1 is transformed into a multiobjective problem, by splitting the objective function (eq. 2) in two objective functions: one for measuring the NN performance, and another one for measuring the complexity of the NN:

$$J_1 = \frac{\rho_1 E_{\text{train}} + \rho_2 E_{\text{gen}}}{\rho_1 + \rho_2} , \tag{5}$$

$$J_2 = C_1 e^{(n_{\text{neurons}})^2} + C_2(n_{\text{epochs}}) + 1 \ . \tag{6}$$

Figure 2 illustrating the operation of NSGA-II implemented for the optimization the parameters of the neural network. Initially a combined population $R_g = P_g + Q_g$ is formed, where P_g represents the current population (parents), and Q_g represent the population of descendants (children) created from the application of genetic operators selection, crossover (μ_c) and mutation (μ_m).

For $i \leftarrow 1, N_{\text{individuals}}$
  P_i^0 = RandomSolution
  Evaluate the two objective functions J_1 *and* J_2
  Train the NN for P_i^0
EndFor
Obtain the non-dominated fronts
P^0 = NonDominatingSortingP^0
$g = 0$
While $g < g_{\max}$
  $P^g_{\text{intermediate}}$ = BinaryTournamentSelectionP^g
  $P^g_{\text{offspring}}$ = Crossover&Mutation$P^g_{\text{intermediate}}$
  Evaluate the two objective functions J_1 *and* J_2
  For $i \leftarrow 1, N_{\text{individuals}}$
    Train the NN for $P^g_{\text{offspring}_i}$
  EndFor
  $P^g_{\text{union}} = P^g \cup P^g_{\text{offspring}}$
  Obtain the non-dominated fronts
  P^g_{union} = NonDominatingSortingP^g_{union}
  Select the best solutions from the sorted union population
  P^{g+1} = SelectBestSolutionsP^g_{union}
  $g = g + 1$
EndWhile
Return P^g

Figure 2: NSGA-II Pseudocode.

4. Climate Precipitation Prediction in Southeast of Brazil

The Southeast region of Brazil was selected to conduct climate prediction precipitation in the seasonal scale. The geographic coordinates delimiting the area considered are between longitudes 52W, 38W and latitude 25S, 15S, as illustrated in Figure 3 (red square on the map of Brazil).

The Southeast region is characterized by the performance of weather systems that combine tropical systems properties with typical mid-latitude systems. During the months of greater convective activity, South Atlantic

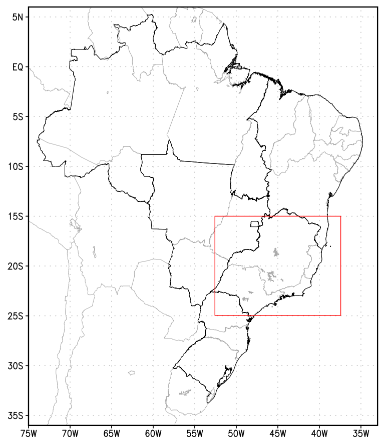


Figure 3: Study Area: Southeast

convergence Zone (SACZ) is one of the phenomena that influence rainfall regime [11].

The Southeast region has low predictability. Therefore, a prediction about the beginning of a wet or dry season is a difficult target to achieve. The Center for Weather Forecasts and Climate Studies (CPTEC) and the world meteorological centers of the USA (National Center for Atmospheric Research (NCAR), National Centers for Environmental Prediction (NCEP), Center for Ocean, Land and Atmosphere Studies (COLA)), Europe (European Centre for Medium-Range Weather Forecasts (ECMWF)), and among others, recognize the complexity in predicting seasonal climate in the Southeast region of Brazil. The forecasts show a low degree of reliability, especially for rain during intermediate seasons like spring and autumn [8].

4.1 - Meteorological data

The data collected for precipitation prediction experiments consists of monthly means from January 1990 to February 2015 (302 months). The data was downloaded from the reanalysis data repository from NCEP/NCAR). The global data reanalysis grid uses a set with horizontal resolution of $2.5^o \times 2.5^o$ (latitude $\times$ longitude). The available variables are presented: temperature (temp), zonal wind components (v) at levels 300, 500, 850 hPa, meridional wind components (u) at levels 300, 500, 850 hPa, surface pressure (spres), specific humidity (shum) and precipitation (prec).

The training data subset was formed with data from January 1990 up to December 2010, this subset was used to derive a NN model. The period from January to December of 2011 was used for the cross validation test, and the period from January 2012 up to February 2015 was used for the generalization.

5. Experiments

Here, four models of networks were constructed: first model, was generated by expert (NN:Expert); second model, was generated using all available variables in the database, with the optimization parameters by using the metaheuristic MPCA (NN:MPCA); third model, containing a reduced number of variables indicated by RST, and self-configuring of parameters by MPCA (NN:RST); and four model, was conducted using all available variables in the database, with the optimization parameters by using multi-objective strategy by NSGA-II (NN:NSGA).

The RST was used to extract relevant information from meteorological data to perform climate prediction, from a reduced data set. The main parameter is the fraction of accuracy, the value of this work was 0.7, that is, all attributes that belong to the joint approximate candidates are reduced, which has at least 70% attendance in discernment function. These are presented as input of the predictive model based on neural networks.

5.1 - Data mining by RST

The variables that form the products has a presence greater than 70% in the discernibility function. It can be noticed that only 5 variables out of 10, were considered. The variables selected by RST are: *temp* (78%), $u850$ (79%), $v300$ (74%), $v500$ (80%) and *shum* (89%). These variables selected were then used to derive the forecasting models for Southeast region of Brazil.

5.2 - Parameter optimization of ANN

Both meta-heuristics are stochastic approaches. Several realizations for both optimizers are executed to find representative solutions. There is a set of parameters to be defined to use the algorithms MPCA and NSGA-II. Table 2 shows the parameters used in our numerical experiments.

The parameters corresponding to the best and worst solutions found after a predetermined number of consecutive evaluations of the objective function (NFE) are show in the Table 3. It is observed that the results obtained by metaheurist NSGA-II, obtained a lower complexity compared to the others approaches.

5.3 - Results

Figure 4(a) shows the reanalysis precipitation field for the Autumn season. Figures 4(b), 4(c), 4(d) and 4(e) show climate prediction results with

Table 2: Control parameters

MPCA	NSGA-II
NFE_{max} 40000	$N_{\text{individuals}}$ 200
$N_{\text{processors}}$ 240	g_{max} 200
$N_{\text{particles}}$ 1	*pool* 100
$N_{\text{blackboard}}$ 100	*tour* 2
IL 0.8	μ_c 20
SL 1.2	μ_m 20

Table 3: Parameters for precipitation prediction

Method	Statistic	n_{neurons}	$f_{\text{activation}}$	η	α	MSE
Expert	–	15	1	0.4	0.9	2.32E-02
	Best	8	1	0.1718	0.2201	3.01E-02
MPCA	Worst	5	2	0.8918	0.0309	3.09E-02
	Mean	6	1	0.5635	0.3457	3.07E-02
	Best	5	2	0.7016	0.9810	2.05E-02
NSGA-II	Worst	5	2	0.5390	0.7462	2.22E-02
	Mean	5	2	0.6163	0.5977	2.20E-02

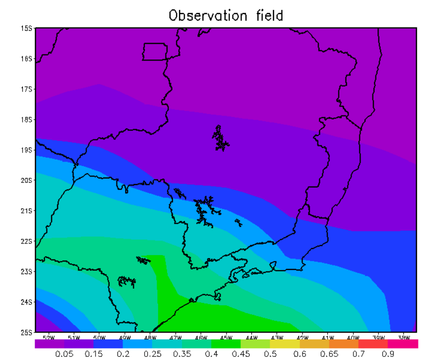
the networks: NN:Expert, NN:MPCA, NN:NSGA and NN:RST, respectively.

From the results shown in Figure 4, the predictions obtained with NN:RST, NN:MPCA, and NN:NSGA identified precipitation patterns similar to the observation field (reanalysis). However, making a qualitative analysis, it is clear that the output produced with the NN:RST model, was detected precipitation levels, especially in São Paulo, an area where the other models did not identify the occurrence precipitation.

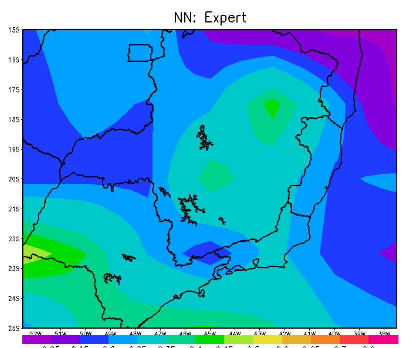
6. Conclusions

This research proposes a methodology that addresses the use of metaheuristics in determining optimal parameters for supervised neural network architectures. In this context, the problem definition of an optimal architecture is formulated as an optimization problem. Two different strategies were applied to the optimization: a mono-objective formulation and a multiobjective formulation. The mono-objective optimization was solved by the MPCA and the NSGA-II was used for multi-objective optimization.

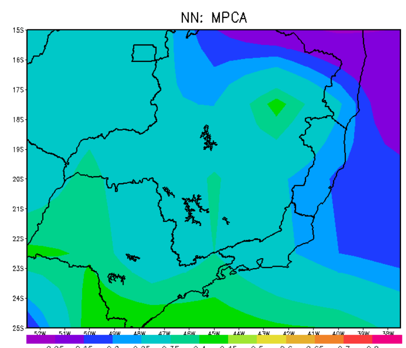
The results for the seasonal forecasts of precipitation using the dimensionality reduction of meteorological data, clearly showed the best results



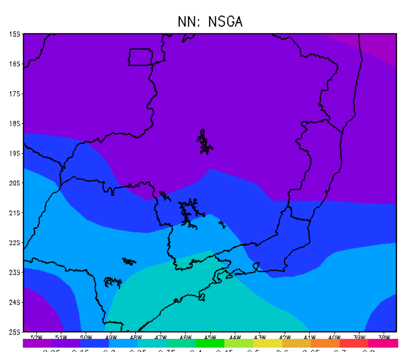
(a) Reanalysis precipitation field NCEP/NCAR: Autumn in Southeast



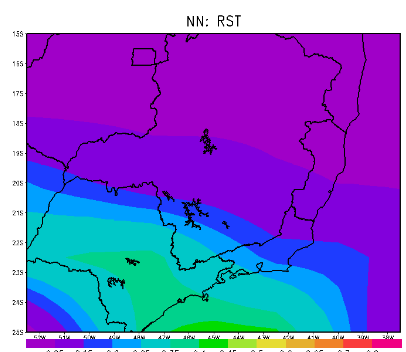
(b) NN:Expert



(c) NN:MPCA



(d) NN:NSGA



(e) NN:RST

Figure 4: Precipitation prediction seasonal for the autumn

in relation to the networks supplied with all the variables available in the database (see Fig.4(e)).

From the results, it has been shown that the use of metaheuristic to find an acceptable topology can be advantageous in relation to an expert. Another advantage is that the automatic strategy discards the need for an expert in neural networks making use of neural networks accessible to a larger audience.

References

[1] Anochi, J.A., Campos Velho, H.F. Optimization of feedforward neural network by Multiple Particle Collision Algorithm, Foundations of Computational Intelligence, IEEE Symposium on. IEEE, 128-134, 2014.

[2] Carvalho, A., Ramos, M., Chaves, A. Metaheuristics for the feedforward artificial neural network (ANN) architecture optimization problem. Neural Comput Applic, 20:1273-1284, 2011.

[3] Benardos, P., Vosniakos, G.C. Optimizing feedforward artificial neural network architecture. Engineering Applications of Artificial Intelligence, Elsevier, 20(3):365-382, 2007.

[4] Deb, K., Pratap, A., Agarwal, S., Meyarivan, T.A. fast and elitist multiobjective genetic algorithm: NSGA-II, Evolutionary Computation, IEEE, 6(2):182-197, 2002.

[5] Dempster, A. Upper and lower probabilities induced by a multivalued mapping. The annals of mathematical statistics, JSTOR, 325-339, 1967.

[6] Komorowski, J., Øhrn, A. Modelling prognostic power of cardiac tests using rough sets. Artificial intelligence in medicine, 15(2):167-191, 1999.

[7] Luz, E., Becceneri, J., Campos Velho, H.F. A new multi-particle collision algorithm for otimization in a high-performance environment, Journal of Computacional Interdisciplinary Sciences, 1:3-10, 2008.

[8] Marengo, J.A., Alvez, L.M., Castro, C.A., Mendes, D. Incio da estao chuvosa nas regies SE/CO/S/NE do Brasil. Centro de Previso de Tempo e Estudos Climticos (CPTEC/INPE), 2015, Available in: http://energia1.cptec.inpe.br/regiaoafetadas/pt.

[9] Metropolis, N., Resenbluth, A., Rosenbluth, M., Teller, A.H. Equation of state calculations by fast computing machines, The Journal of Chemical Physics, 6(21):1087-1092, 1953.

[10] Pawlak, Z. Rough sets. International Journal of Computer & Information Sciences, Springer, 11(5):341-356, 1982.

[11] Quadro, M., Machado, L., Calbete, S., Batista, N., Oliveira, G. Climanlise: Boletim de Monitoramento e Anlise Climtica. Centro de Previso de Tempo e Estudos Climticos (CPTEC/INPE), 9(1), 1996. Available in: http://climanalise.cptec.inpe.br/ rcliманl/boletim/cliesp10a/chuesp.html.

[12] Sacco, W., Oliveira, C. A new stochastic optimization algorithm based on a particle collision metaheuristic. Proceedings of 6th World Congress of Structural and Multidisciplinary Optimization, WCSMO, 2005.

[13] Zadeh, L. Fuzzy sets. Information and Control, Elsevier, 8:338-353, 1965.

[14] Zadeh, L. Fuzzy sets as a basis for a theory of possibility. Fuzzy sets and systems, Elsevier, 1:3-28, 1978.