

Climate Precipitation Prediction by Neural Network

Juliana Aparecida Anochi¹, Haroldo Fraga de Campos Velho²

1. *Applied Computing Graduate Program, National Institute for Space Research (INPE), Brazil.*

2. *Associated Laboratory for Computing and Applied Mathematics, National Institute for Space Research (INPE).*

Received: December 08, 2014 / Accepted: January 10, 2015 / Published: May 25, 2015.

Abstract: In this work a neural network model for climate forecasting is presented. The model is built by training a neural network with available reanalysis data. In order to assess the model, the development methodology considers the use of data reduction strategies that eliminate data redundancy thus reducing the complexity of the models. The results presented in this paper considered the use of Rough Sets Theory principles in extracting relevant information from the available data to achieve the reduction of redundancy among the variables used for forecasting purposes. The paper presents results of climate prediction made with the use of the neural network based model. The results obtained in the conducted experiments show the effectiveness of the methodology, presenting estimates similar to observations.

Key words: Climate Prediction, Neural Networks, Rough Sets Theory

1. Introduction

The development of regional climate models from data considers the hypothesis that it is possible to extract information from historical data on the behavior of climatic conditions. Accordingly, development methodology needs to have a large number of data to ensure that the model considers a wide range of situations. However, despite able to ensure, in principle, greater robustness to the derived models, the handling of large volumes of data may require much computational. This problem is even more complicated when dealing with weather forecasting tasks that require higher accuracy once it deals directly with every-day life of the population.

In order to cope with the demand for computing performance, it is necessary to assemble supercomputers or clusters of powerful computers that may work together to produce numerical results that are inputs for forecasting tasks. Nevertheless, it is not only the computing power that counts in a climate forecasting task. In these regards, it would be very

profitable if computational models were constructed based on the existing knowledge. This would permit forecasters to run the climate model for a several different initial conditions thus leading to a variety of different scenarios to be analyzed to support their decision on the prediction of the state of the atmosphere. In order to achieve a computational model that may satisfy the stated conditions, we proposed the use of artificial neural networks to be trained with existing climate data.

In this paper considered the use of Rough Sets Theory principles in extracting relevant information from the available data to achieve the reduction among the variables used for forecasting purposes. The data reduction approach allows the derivation of smaller data sets from the resulting reducts for the training phase of the neural network without losing data expressiveness for forecasting purposes

The advantage of the methodology presented in this paper is the possibility to derive climate forecasting models from simplified data sets gather from observations and to have a means to make prediction using simplified non-complex computational models based on neural networks. An additional advantage of

Corresponding author: Juliana Aparecida Anochi, Applied Computing Graduate Program, National Institute for Space Research (INPE), Brazil. E-mail: juliana.anochi@gmail.com.

the approach is the possibility to derive a model without a need for deep understanding of the physics related to atmosphere processes.

Artificial Neural Networks (ANN) have emerged as excellent tools for deriving data oriented models because of their inherent characteristic of plasticity that permits the adaptation of the learning task when data is provided. In this regards, the use of an ANN is adequate to derive the forecasting model proposed in this paper. For the problem of building prediction models, as proposed in this paper, are used Multilayer Perceptron and models of recurrent Elman and Jordan [2].

2. Rough Sets 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 [7]. The RST is based on the similarities among objects measured by an indiscernibility relation, which establish that a set of objects are similar (indiscernible) if they hold the same values for all of their attributes. Rough sets have been proposed for a very wide variety of applications. In particular, the rough set approach seems to be important for Artificial Intelligence, especially in machine learning, knowledge discovery, data mining, expert systems, approximate reasoning and pattern recognition.

In this paper, the RST was used to extract relevant information from meteorological data to perform climate prediction, from a reduced data set.

2.1 Information Systems

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]. More formally, it is a pair $SI = (U, A)$, where U is non-empty finite set of *objects* called the *universe* and A is a non-empty finite set of *attributes*, 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*.

2.2 Indiscernibility

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 [5]. This is equivalence relation that 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. For given information system $SI = (U, A)$ for any set $B \subseteq A$ there exists an equivalence relation $IND_A(B)$, as defined in Equation 1, that establish two objects x and x' in U are indiscernible from each other by attributes from B .

$$IND_A(B) = \{(x, x') \in U^2 \mid \forall a \in B, a(x) = a(x')\} \quad (1)$$

2.3 Attribute Reduction

The attribute reduction procedure is performed by the discernibility function $f_A = (B)$ derived from the discernibility matrix which is a symmetric matrix constructed by comparing the attribute values that discern the objects. The attribute representing discernible values are inserted into the matrix. Each entry in the matrix consists of a set of attributes that distinguish a pair of objects x_i and x_j as it is stated in

$$M_{i,j} = \{a \in B \mid a(x_i) \neq a(x_j)\} \quad (2)$$

where $1 \leq i, j \leq n$ and $n = |U / IND_A(B)|$.

The discernibility function $f_A = (B)$ for a subset B is constructed by concatenating the subsets $M_{i,j} = \{a^* \mid a \in M_{i,j}\}$ of attributes in each position (cell) of the discernibility matrix M , through a Boolean function of m variables a_1^*, \dots, a_m^* (corresponding to the attributes) defined as below. The function determines the

minimum set of attributes that distinguish any class among the existing classes and it is defined as [7]:

$$f(a_1^*, \dots, a_m^*) = \wedge \{ \vee C_{ij}^* \mid 1 \leq j \leq i \leq n, \quad C_{ij} = \phi \} \quad (3)$$

3. Artificial Neural Network

Artificial Neural Networks are computational techniques that present a mathematical model inspired on the neural structure of biological organisms, acquiring knowledge through experience. The network has parallel and distributed processing units or neurons, which calculates certain mathematical functions, typically nonlinear. These neurons can be divided into one or more hidden layers interconnected by synaptic weights (connections), which store the knowledge represented in the model, and serve to balance the input received by each network neuron [2].

The artificial neuron model proposed by McCulloch and Pitts in 1943, interprets the functioning of the neuron as a simple binary circuit that combines multiple inputs and one output signal. The mathematical description resulted in a model with n input terminals representing the dendrites, and only one output simulating the axon. In the mathematical terms, a neuron k could be described considering the following pair of equations [2]:

$$v_k = \sum_{j=1}^n w_{kj} x_j \quad (4)$$

$$y_k = \varphi(v_k + b_k) \quad (5)$$

where x_1, x_2, \dots, x_n are the input signals; w_{kj} are the synaptic weights of neuron k computed during the learning phase; v_k is the linear combination among the input signals; b_k is the bias; φ is the activation function; and y_k is the output signal of the neuron. The use of bias b_k has the effect of applying an affine transformation to the output v_k of the linear combiner in the model.

The property of primary significance for a neural

network is the ability to learn from the environment, improving its performance through the learning. A neural network learns from its environment through an interactive process of adjustments applied to its synaptic weights and bias levels [2]. In the context of NN, the learning process can be defined as a set of well defined rules for solving a specific problem.

3.1. Multilayer Perceptron

The architectures of multilayer perceptron neural models are the most used and known. Typically, this architecture consists of a set of units forming an input layer, one or more intermediate (or hidden) layer(s) of computational units, and an output layer. It is a supervised neural network.

A very popular algorithm for training MLP-NN is called backpropagation of error. The backpropagation learning error consists of two steps through the different layers of the network: forward and backward steps. For the forward step, an activity pattern (input vector) is applied through the nodes of the network, and its effect propagates on the entire the network, layer by layer. Finally, a set of outputs is produced as the actual response of the network. During the backward step, the synaptic weights are all adjusted in accordance with an error correction rule. The response of the network is subtracted from a desired output to produce an error signal. Figure 1 shows the architecture of a multilayer perceptron.

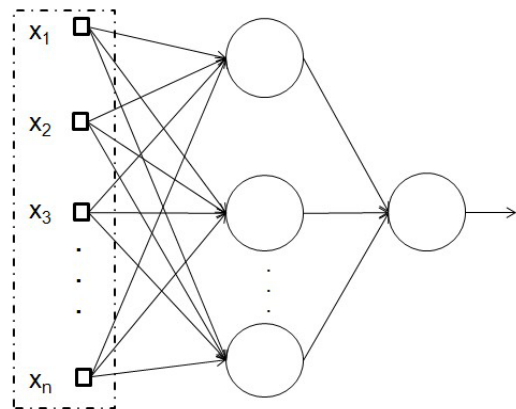


Fig. 1 Network multilayer perceptron with an input layer, a hidden layer, and an output layer.

3.2. Recurrent Neural Network

The recurrent networks are neural networks with one or more feedback loops. Given a multilayer perceptron as the basic building block, the application of global feedback can take a variety of forms. We may have feedback from the output neurons to the input layer. Another possibility is the link among the hidden neurons with the input layer. When the MLP has two or more hidden layers, the possible forms of global feedback expand even further [3].

The recurrent networks Elman and Jordan consists of four layers: an input layer, composed of storage neurons that receive an external signal to propagate it, without modification; a hidden layer where activation functions may be linear or non-linear; a context layer is used to store the activation of neurons; the intermediate layer (it can be used with time delay), and an output layer, neurons whose outputs are linear sums of their respective input signals).

4. Results and Discussion

In the dimensionality reduction process the relevant attributes are those that mostly occur in the data, in terms of the indiscernibility relation. The data consists of monthly means from January 2000 to December 2009. The analysis region comprises a selected subregion with coordinates (Lat 22.5S, 35S) to (Lon 60W, 45W). The data was downloaded from the reanalysis data repository from National Center for Environmental Prediction & National Center for Atmospheric Research (NCEP/NCAR). The world is divided into parallels (latitudes) and meridians (longitude). The global data reanalysis grid uses a set with horizontal resolution of $2.5^\circ \times 2.5^\circ$ (latitude x longitude) [4]. In order to evaluate the performance of the models, the mean square error is used:

$$Error = \frac{1}{N} \sum_{k=1}^N (y_k - d_k)^2 \quad (6)$$

where N is the number of patterns in the data set, y_k is the true observational value, and d_k is the estimation

computed by neural model.

The available variables are: air temperature (airt), zonal wind components at vertical levels: 300hPa (v300), 500hPa (v500) and 850hPa (v850), meridional wind components at vertical levels: 300hPa (u300), 500hPa (u500) and 850hPa (u850), surface pressure (spres), specific humidity (shum) and precipitation (prec).

The variables that were reduced by the use of RST are: airt (77%), u300 (79%), u850 (80%), v300 (77%), v500 (79%), v850 (70%) and spres (78%). It is to be noticed that the variables that form the reducts have a presence greater than 70% in the discernibility function.

It is observed that of 7 variables out of 10, were considered relevant for South. Figure 2 shows the results for season winter of 2007 in the South region of Brazil. Figure 2-a is the observed precipitation in winter 2007; Figure 2-b corresponds to network Jordan trained with all the available variables in the database; Figure 2-c is the estimation produced by the Elman trained with all the variables; and Figure 2-d corresponds to the estimation produced by the MLP network trained with all the available variables.

From the results, the Elman network trained with all the available variables in database presented the best performance than the others networks.

Figure 3-a is the observed precipitation in winter 2007; Figure 3-b corresponds to the result produced by the recurrent network Jordan trained with the reduced set of variables processed by rough set theory; Figure 3-c shows the result obtained by the network recurrent Elman that was trained with the reduced set of variables processed by RST; and Figure 3-d corresponds to the estimation produced by the MLP network trained with the reduced set by RST.

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. Indeed, both models ANN-Elman and ANN-MLP computed by RST show a high fidelity with real dynamics.

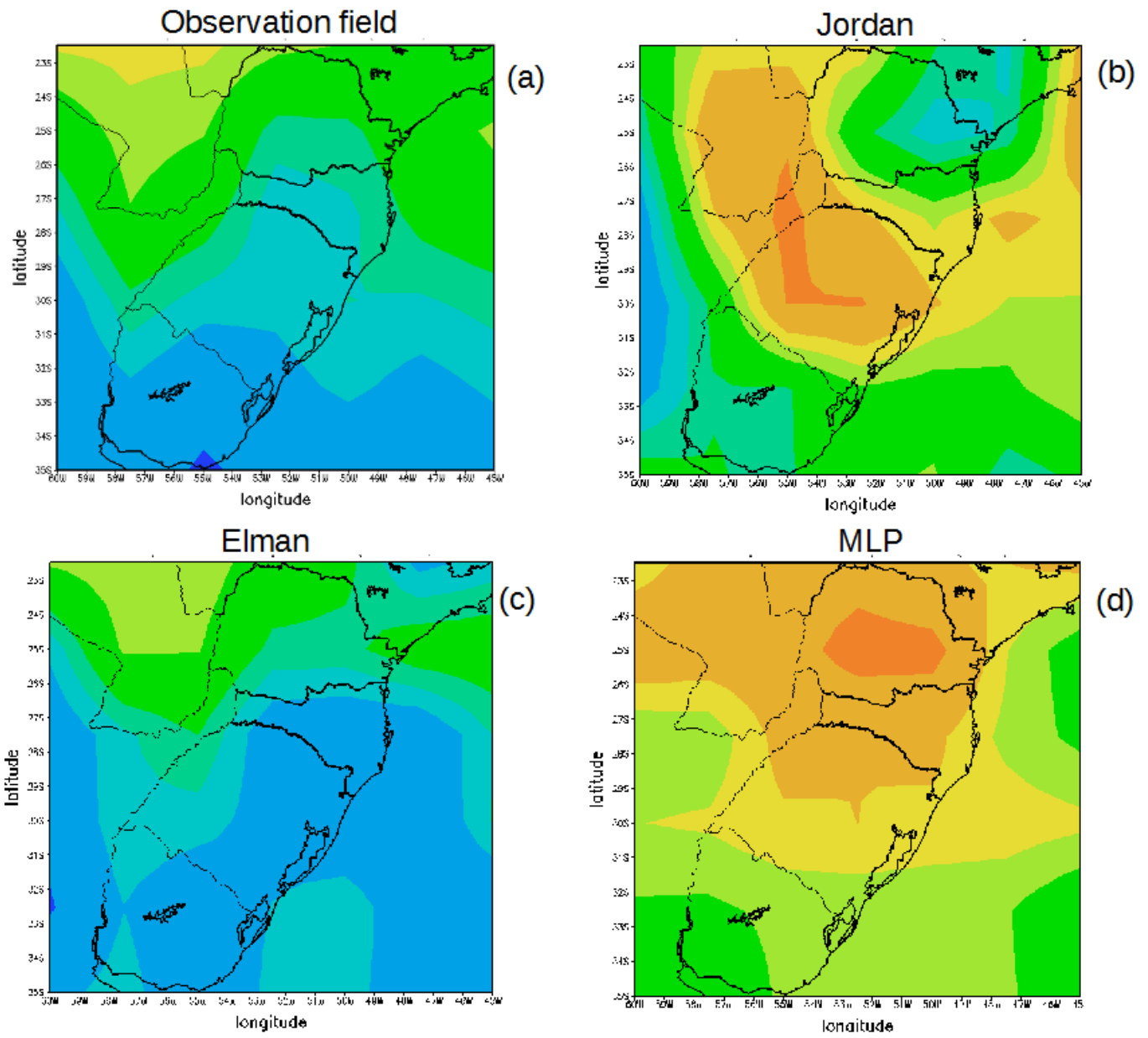


Fig. 2 (a) observed precipitation; (b) ANN-Jordan; (c) ANN-Elman; and (d) MLP network.

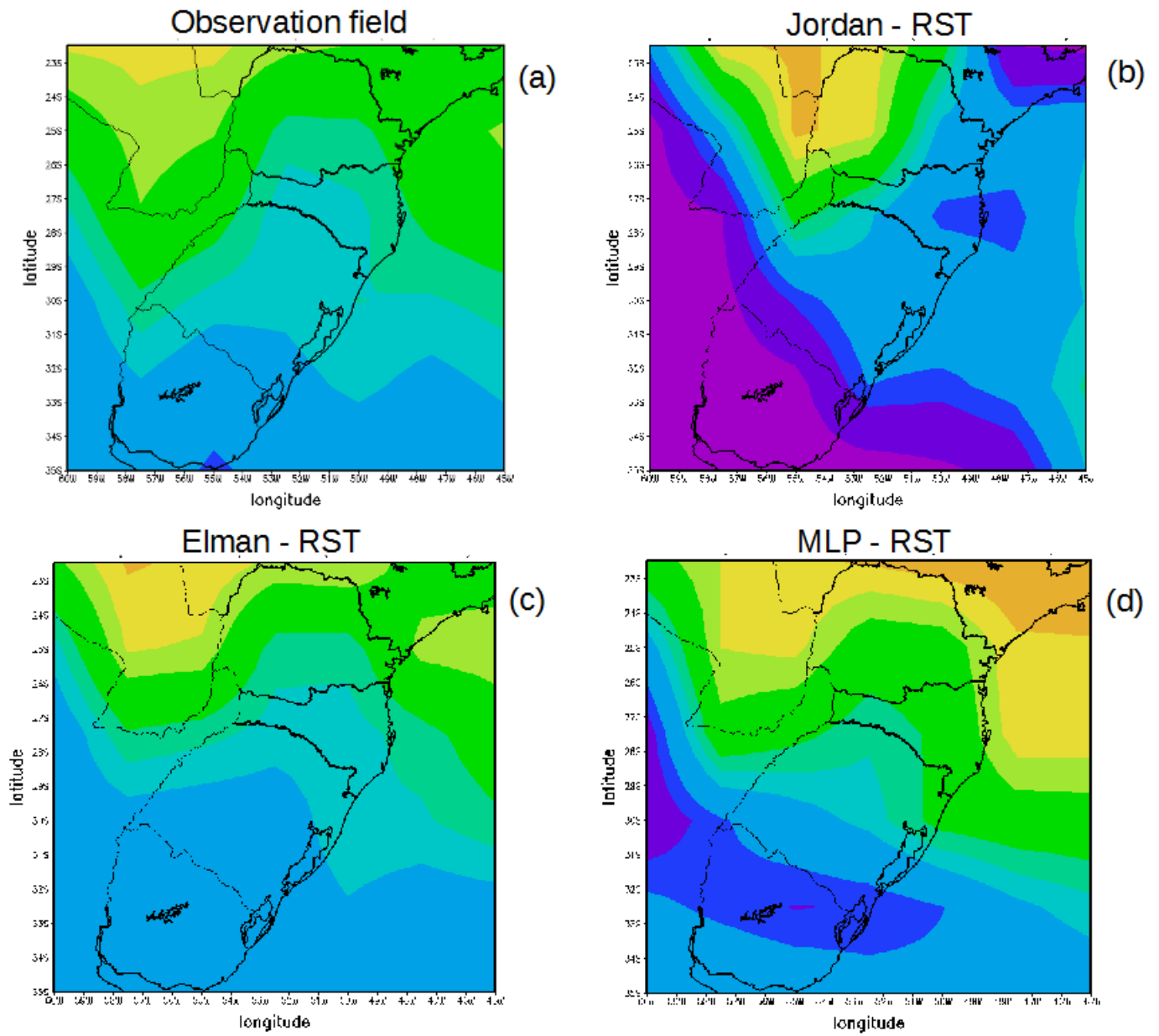


Fig. 3 (a) Observed precipitation; (b) NN-Jordan; (c) NN-Elman; and (d) NN- MLP.

In addition, a significant performance was obtained when the training data was processed to eliminate the existing redundancy in the data. The results show that it is worth to search for hidden information in the data in a way that less (computational) efforts are necessary to construct efficient forecasting models.

5. Conclusions

In this paper a methodology was used for data reduction, applied in the study of seasonal weather patterns, using the theory of rough sets (data reduction) and neural networks (forecast) in the derivation of predictive models in order to estimate the state future precipitation variable, in south Brazil.

The proposed methodology works well with Elman-ANN using RST. The data reduction approach allows the derivation of smaller data sets derived from the resulting reducts for the training phase of the neural network without losing data expressiveness for forecasting purposes. Thus, the RST technique used in the data reduction process allows the identification of relevant information of the data for climate prediction. In addition, it is a technique that approaches the

problem of dealing with huge amounts of data which is a characteristic of the processes in meteorology.

Acknowledgment

The authors thank CNPq and CAPES for the physical and financial support provided to this research.

References

- [1] L. Elman, Finding Structure in Time, *Cognitive Science*. 14 (1990) 179-211.
- [2] S. Haykin, 1998, *Neural networks principles and practices*, Prentice Hall.
- [3] L. Fausett, 1993, *Fundamentals of Neural Networks: Architectures, Algorithms And Applications*, Prentice-Hall.
- [4] E. Kalnay, 2003, *Atmospheric modeling, data assimilation, and predictability*. Cambridge university press.
- [5] J. Komorowski, Z. Pawlak, L. Polkowski, A. Skowron, Rough sets: A tutorial. Rough fuzzy hybridization: A new trend in decision-making. 1999, pp. 3-98.
- [6] J. Komorowski. A. Ohrn, Modelling prognostic power of cardiac tests using rough sets, *Artificial Intelligence in Medicine*, 15 (1999) 167-191.
- [7] Z. Pawlak, Rough sets, *International Journal of Computer and Information Sciences*, 1(1982) 341-356.