Development of information and analytical tools based on adaptive classifier Cascade ARTMAP

I L Kashirina, K A Fedutinov and Yu V Bondarenko

Department of Applied Mathematics, Informatics and Mechanics, Voronezh State University, 1 Universitetskaya pl., Voronezh, 394006, Russia

E-mail: kash.irina@mail.ru

Abstract. The article deals with methods of teaching incremental classifiers capable to adapt to incoming online streaming data, using the example of environmental monitoring systems. It is proposed to use Cascade ARTMAP neural network architecture, compatible with symbolic representation based on IF-THEN rules. The knowledge gained during Cascade ARTMAP network training can be transformed into a compact set of decisive rules for classification of source data that can be easily interpreted by experts in the field of environmental assessment.

1. Introduction

Modern environmental monitoring systems are complex information systems of observation, assessment and forecast of changes in the state of environmental conditions, as well as development of management decisions (Figure 1). In connection with the widespread introduction of methods of artificial intelligence in the solution of real problems, the creation of automated decision support systems based on machine learning methods, allowing the effective use of monitoring data, becomes relevant today.

Figure 1. Environmental Monitoring Scheme.

This article deals with the task of improving the efficiency of environmental monitoring through the development of incremental neural network classifiers and their introduction into the decision support system within an automated information system. Neural networks can work well in the
conditions of the incomplete input information, and also in the conditions of noise and hindrances that often is characteristic for monitoring indicators, hence, they can raise efficiency and accuracy of an estimation of a current condition of environment, and, hence, efficiency of accepted management decisions.

Incremental learning allows the adaptation of classification algorithms in real time, which makes it possible to process the incoming information on the state of the environment (air temperature, information on pollutant emissions, wind direction, wind speed, etc.) quickly. The research of existing neural network (NN) architectures [5] has shown the potential of using the theory of adaptive resonance (ART) in environmental monitoring systems. Networks of ART family allow to realize qualities important for ecological monitoring systems: stability (preservation of accumulated knowledge) and flexibility (continuous updating of knowledge in the process of incremental learning). In addition, networks of architecture Cascade ARTMAP provide a unique opportunity to display the a priori experience of environmental experts in the structure of clusters to form the initial knowledge base, which is completed and adjusted in the process of network operation [6].

![Figure 2. Basic module of ART family networks.](image)

2. Materials and methods
Adaptive Resonance Theory (ART) is the basic theory of the family of neural networks, which in the adaptive learning mode are able to form stable clusters in response to arbitrary incoming sequences of input data on the basis of self-organization. By dynamically creating clusters, ART family networks can adjust their size and number depending on the difficulty and complexity of the input data set. As of today, this family includes more than 10 different architectures, each of which contains one or two basic modules presented in Figure 2. The algorithm for the functioning of such a basic module is based on the search of the neuron containing the cluster prototype closest to the incoming input vector.
The search for the prototype is done in two stages. The first stage calculates the activation levels for all neurons-prototypes. The neuron with the highest activation value is marked as the "winner". At the second stage the "winner" is checked for resonance with an input vector. If the resonance value exceeds the specified threshold value, weights of the found "winner" are corrected to approach the current input vector. If the resonance has not exceeded the threshold value, the next "winner" is searched for and the one found earlier is excluded from consideration. If a prototype similar to the input vector is not found, a new prototype is created and its weights are selected equal to the current input. The characteristic feature of ART family networks is that they increase in size if they do not find in memory a prototype similar to the input vector. If the prototype is found, the network does not increase in size and the weights of the "winner" neuron are adjusted.

Figure 3. ARTMAP Network Architecture.

Like other networks in the ART family, the ARTMAP network is an online real-time system that does not require a separate pre-learning phase, i.e. the system operates according to the current input environment while learning.

The ARTMAP network consists of two basic ART modules: ARTa and ARTb, combined using an associative \( F^{ab} \) memory layer, as shown in the Figure. 3 and based on the principles of fuzzy logic. Here ARTa is the module for processing the input vector A, ARTb is the module for processing the output vector B; \( F^a_1 \) and \( F^b_1 \) are the comparison layers, \( F^a_2 \) and \( F^b_2 \) are the recognition layers, bound together by sets of weighting coefficients \( w^a_j \) and \( w^b_j \) respectively.

Each neuron of recognition layers \( F^a_2 \) and \( F^b_2 \) represents some recognition category, i.e. a cluster of similar input images characterized by a common vector-prototype stored in the vector of weight coefficients of this neuron. Binary weight \( w^a_{jk} \) connects layer \( F^a_2 \) with layer \( F^b_2 \), which implements the mechanism of self-regulation of connections between formed categories of input and output vectors. The ARTMAP network deals with the problem of classification, and each class can correspond to several categories (clusters).

In the ARTMAP network, the generalization of input images is done not by their algebraic averaging, but by applying the fuzzy operation "AND". As a result of calculation of this operation, the values of weights are reduced from the maximum value to a certain stable state close to the values of some group of input vectors. This is a stable state and will be a prototype of the cluster. In order to prevent such a decrease from degenerating the prototypes, the ARTMAP network applies complementary coding to input and output vectors: an M-dimensional vector is converted into a 2M-
dimensional vector by adding M components that allow to save amplitude information. In this way, the input images of the ARTMAP network look like: \( A = (a, a^c) \), where \( a_i = 1 - a_i^c \); output images have the form \( B = (b, b^c) \), where \( b_i^c = 1 - b_i \).

The Cascade ARTMAP network aimed at extracting rules is a modification of the ARTMAP network. It combines layer \( F_1^a \) of input attributes and layer \( F_1^b \) of output attributes in the sense that both \( F_1^a \) and \( F_1^b \) will handle input, output and intermediate attributes. Let’s consider two rules, which form a simple two-level rule cascade:

Rule 1: IF A and B THEN C,
Rule 2: IF C and D THEN E,

where A, B and D - input attributes; C - intermediate attributes; and E - output attributes. All attributes (A, B, C, D and E) are represented both in layer \( F_1^a \), and in layer \( F_1^b \) (Figure 4).

For rule 1, the category of layer \( F_2^a \) is used to memorize inputs A and B, and is associated with the category of layer \( F_2^b \) that predicts output C. Similarly to rule 2, layer category \( F_2^a \) is used to memorize inputs C and D, and is associated with layer category \( F_2^b \), which predicts output E.

![Figure 4. Presentation of a cascade of simple rules in Cascade ARTMAP network.](image)

Thus, the structure of Cascade ARTMAP network is compatible with the representation of knowledge in expert systems, and the rules of fuzzy logic output can be transformed into specific recognition categories. Initialization of Cascade ARTMAP with the a priori known rules of fuzzy logic output forms the initial structure of the neural network and helps accelerate the learning process and accuracy of forecasting, because it generates clusters that may not be currently represented in the source data.

The process of adding fuzzy logic output rules is performed in two stages: 1) attribute names are extracted from all the rules, and a \( Z \) table is created with a full list of all the attributes \( (z_1, z_2, ..., z_M) \); 2) based on data from the attribute table, each rule is converted into two vectors A and B, which are used as input vectors for ARTa and ARTb modules, respectively.

Thus, the rule “IF \( x_1, x_2, ..., x_M \) THEN \( y_1, y_2, ..., y_M \)” taking into account complementary coding will correspond to the binary input vector \( A = (a_1, a_2, ..., a_M, a_1^c, a_2^c, ..., a_M^c) \) and binary output vector \( B = (b_1, b_2, ..., b_M, b_1^c, b_2^c, ..., b_M^c) \), where

- \( a_i = 1, a_i^c = 0 \), if \( x_i = z_i \);
- \( a_i = 0, a_i^c = 1 \), if \( x_i = \bar{z}_i \);
- \( a_i = 0, a_i^c = 0 \), if the \( i^{th} \) attribute of the table is not presented in the hypothesis of the rule.
Analogous:

- \( b_i = 1, b_c^j = 0 \), if \( y_i = z_i \);
- \( b_i = 0, b^c_i = 1 \), if \( y_i = \overline{z_i} \) and
- \( b_i = 0, b^c_i = 0 \), if the \( i \)-th attribute of the table is not represented in the rule output.

To learn the Cascade ARTMAP network, a reverse tracking algorithm is used that identifies all rules (layer categories \( F^a_2 \) and \( F^b_2 \)), that are responsible for obtaining the current output value. It means that if, for example, the current input vector activated the \( j \)-category from layer \( F^a_2 \), the algorithm identifies the previous set of categories that led to the selection of the \( j \)-category. The reverse tracking in this case is in the direction of \( F^a_2 \) → \( F^b_1 \) → \( F^b_2 \) → \( F^{ab} \) → \( F^a_2 \). For example, in Figure 4, the algorithm will trace the category \( j \) of layer \( F^a_2 \) to its predecessors C and D in layer \( F^a_1 \). It then checks that C is an intermediate attribute activated in \( F^b_2 \) and finally tracks the category \( j \) in layer \( F^b_2 \). Tracking stops at \( j \), because all the predecessors of this category are input attributes. That is, the predecessor of category \( j \) of layer \( F^a_2 \) is category \( j \), so the learning algorithm will correct not only the weights of the neuron \( j \), but also the weight coefficients of the neuron \( j \). The algorithm for correcting the weights of the Cascade ARTMAP network is described in detail in [8].

The process of extracting rules from Cascade ARTMAP network is quite simple. As already noted, each neuron in the layer \( F^a_2 \) is a category of recognition of input templates of ARTa module. Through the intermediate layer \( F^{ab} \) each such neuron is associated with the ARTb module category in layer \( F^b_2 \), which, in turn, encodes the result of the classification. Trained weight vectors of layer \( F^a_2 \), memorize a set of rules that link the input data with the results. The total number of rules is equal to the number of neurons of layer \( F^a_2 \), which were formed during the learning.

As a rule, large data sets force Cascade ARTMAP to generate too many rules for their practical application. The task of extracting rules is to select a small set of rules for well-predictable categories and describe them in a clear form. To evaluate each category, a validation factor is calculated which determines the accuracy of the rule corresponding to that category. Removing the low-fidelity recognition categories created from atypical examples results in smaller networks. The fidelity factor is calculated for each category of layer \( F^a_2 \) in terms of the frequency of occurrence of this category in the training sample and the accuracy of classification of objects belonging to this category. To estimate the coefficient of accuracy, two indicators are calculated for each \( j \)-category of layer \( F^a_2 \) the fraction \( c_j \) of associated category \( j \) of training vectors and the fraction \( s_j \) of objects correctly classified by category \( j \) (relative to the number of all objects belonging to this category). The node reliability factor \( j \) is calculated as follows: \( k_j = \gamma c_j + (1 - \gamma) s_j \), where \( \gamma \in [0, 1] \) — is a weight coefficient.

3. Results and discussion

To research the possibility of applying the Cascade ARTMAP network as part of environmental monitoring systems, there was developed software in R language using RSNNS library. This library contains the basic implementation of the ARTMAP network [9], by means of which the modified model of Cascade ARTMAP was built.

As the initial data we used environmental monitoring indicators taken in Voronezh region with 500x500 m resolution [6]. The set of processed data made up 206 676 values for each of 17 initial indicators (including geographical coordinates of the point of indicator values collection). The value of each indicator is measured on an integer scale. The data are provided by the Department of Geological Ecology of the Voronezh State University, where a large amount of actual material for the analysis of natural and anthropogenic indicators of the Voronezh region has been accumulated.

To verify the network performance, the experts formulated 18 rules for assessing the environmental condition at each monitoring point based on the values of indicators in 5 classes of environmental risk: "low risk", "medium risk", "increased risk", "high risk". The results are presented in Table 1.
After training the network on some subset of input vectors (0%, 10% or 50% of the input data) the testing procedure was started. The size of the training sample of 0% means that the network was initialized by the specified set of rules, and then the incremental training mode of Cascade ARTMAP was started, whereas the accuracy estimation in the table is given for the whole set of source data. If the size of the training sample is, for example, 10%, it means that the initial structure of the network was formed on the basis of 10% of the source data and the initial values of weights were determined, and the accuracy assessment in the table is given for the remaining 90% of the source data.

Initially, the quality of classification using the Cascade ARTMAP network was checked without using the initial database of classification rules. Further, the network was initialized with a subset of rules from the initial knowledge base, and then trained and tested as in the previous experiment. It follows from Table 1 that Cascade ARTMAP extracts hidden regularities from the data set quite accurately. On average (depending on the hyperparameter settings of the network, in particular, the threshold value for the rule validity factor), 15-25 rules are created in the process of functioning. Initial initialization with the rules significantly increases the network's recognition ability. When Cascade ARTMAP is initialized with a full set of 18 rules, 76.7% of the source data are successfully classified even without the preliminary training stage.

| Network                              | Size of training sample (%) | Rules created (in average) | Classification accuracy (%) |
|--------------------------------------|-----------------------------|---------------------------|----------------------------|
| Cascade ARTMAP without rules         |                             |                           |                            |
| initialization                       | 10                          | 17.4                      | 88.2                       |
|                                      | 50                          | 22.3                      | 98.1                       |
| Cascade ARTMAP with 6 rules          |                             |                           |                            |
| initialization                       | 0                           | 6.4                       | 49.1                       |
|                                      | 10                          | 19.1                      | 96.9                       |
|                                      | 50                          | 23.2                      | 98.9                       |
| Cascade ARTMAP with 12 rules         |                             |                           |                            |
| initialization                       | 0                           | 12.5                      | 64.2                       |
|                                      | 10                          | 18.6                      | 95.4                       |
|                                      | 50                          | 25.3                      | 99.8                       |
| Cascade ARTMAP with 18 rules         |                             |                           |                            |
| initialization                       | 0                           | 18.2                      | 76.7                       |
|                                      | 10                          | 19.7                      | 97.8                       |
|                                      | 50                          | 25.4                      | 100                        |

4. Conclusion
As a result of the research the software allowing to use adaptive classifiers on the basis of neural network architecture Cascade ARTMAP as a part of ecological monitoring systems was developed and tested on the example of Voronezh region. The developed approach allows to investigate spatial interaction of system elements, to reveal leading factors influencing an ecological condition and in an automatic mode to form classification rules in visual and clear format IF - THEN for experts in a subject area. The characteristics of ART networks, in particular, the possibility of continuous learning, when new data coming to the input are used to expand knowledge in the existing model, is a computational paradigm that allows you to form control solutions "on the fly", as new monitoring information is available. The developed algorithms can be used to process streaming data of an arbitrary nature or big data, allowing, without a long-lasting pre-training procedure, to extract knowledge from this data and generate management solutions based on this knowledge.
The proposed information-analytical toolkit is based on modern machine learning methods and can be used in regional systems of the digital economy to study the relationships between the structural components of the main factors of the region’s economic development.

Acknowledgments
The research was supported by the Russian Foundation for Basic Research (RFBR), project No. 19-29-07400.

References
[1] Burkov V N, Novikov D A and Shcheplin A V 2008 Mechanisms of management of ecological and economic systems (FIZMATLIT) pp 244
[2] Tan A H 1997 Cascade ARTMAP: Integrating Neural Computation and Symbolic Knowledge Processing (IEEE Trans.on Neural Networks) 2 vol 8 pp 34-52
[3] Carpenter G A and Grossberg S 2003 Adaptive Resonance Theory (The Handbook of Brain Theory and Neural Networks, Cambridge) pp 87-90
[4] Kashirina I L and Fedutinov K A 2018 Clustering of continuous data flow on the basis of generalized neural network model of ART family (Control Systems and information technologies) 1 pp 33-39.
[5] Carpenter G A, Grossberg S and Reynolds J H 1991 ARTMAP: Supervised real-time learning and classification of nonstationary data by a self-organizing neural network (Neural Networks) 4 pp 565-588.
[6] Kashirina I L and Fedutinov K A 2018 System analysis of monitoring ecological information by means of neural network Fuzzy ART (Actual problems of applied mathematics, Informatics and mechanics proceedings of the International scientific conference) pp 1565-1571.
[7] Kashirina I L and Fedutinov K A 2018 Application of Fuzzy ARTMAP network in intelligent intrusion detection systems (Modeling, optimization and information technologies) 3 vol. 6 pp 243-257.
[8] Kashirina I L and Fedutinov K A 2018 Construction of decisive rules with the help of neural network ARTMAP (Modeling, optimization and information technologies) 3 pp 17-24.
[9] Bergmeir C and Benitez J 2012 RSNNS: Neural Networks using the Stuttgart Neural Network Simulator (Journal of Statistical Software) 46
[10] Nesteruk F G, Tatarinov A Yu and Nesteruk G F 2010 Research of adaptive classifiers as a part of intellectual means of information protection (Optoelectronic information and energy technologies) 2 pp 110-118.