ADAPTIVE NEURO-FUZZY KOHONEN NETWORK WITH VARIABLE FUZZIFIER

Yevgeniy Bodyanskiy, Bogdan Kolchygin, Iryna Pliss

Abstract: The problem of neuro-fuzzy Kohonen network self-learning with fuzzy inference in tasks of clustering in conditions of overlapped classes is considered. The basis of the approach are probabilistic and possibilistic methods of fuzzy clustering. The main distinction of the introduced neuro-fuzzy network is the ability to adjust the values of fuzzifier and synaptic weights in on-line mode, as well as the presence except convenience Kohonen layer an additional layer to calculate the current values of the membership levels. The network characterized of computational simplicity, and is able to adapt to data uncertainty and detect new clusters appearance in real time. The experimental results confirm effectiveness of the approach developed.

Keywords: clustering, neuro-fuzzy network, self-learning algorithm, self-organizing Kohonen map.

ACM Classification Keywords: I.2.6 Artificial Intelligence - Learning - Connectionism and neural nets

Introduction

The problem of multidimensional data clusterization is an important part of exploratory data analysis [Tukey, 1977; Höppner, Klawonn, Kruse, Runkler, 1999], with its goal of retrieval in the analyzed data sets of observations some groups (classes, clusters) that are homogeneous in some sense. Traditionally, the approach to this problem assumes that each observation may belong to only one cluster, although more natural is the situation where the processed vector of features could refer to several classes with different levels of membership (probability, possibility). This situation is the subject of fuzzy cluster analysis [Bezdek, 1981; Gath, Geva, 1989; Höppner, Klawonn, Kruse, Runkler, 1999], which is based on the assumption that the classes of homogeneous data are not separated, but overlap, and each observation can be attributed to a certain level of membership to each cluster, which lies in the range of zero to one [Höppner, Klawonn, Kruse, Runkler, 1999].

Initial information for this task is a sample of observations, formed from N-dimensional feature vectors $x(1), x(2), \dots, x(k), \dots, x(N)$. The result of clustering is segmentation of the original data set into m classes with some level of membership $u_j(k)$ of k-th feature vector $x(k)$ to j-th cluster, $j = 1, 2, \dots, m$.

In this paper we propose computationally simple adaptive procedure for recurrent fuzzy clustering of data processing in real time mode for operating in conditions of a priori uncertainty about the boundaries between classes, and their using for learning of fuzzy self-organizing Kohonen network.

Probabilistic fuzzy clustering

In the class of fuzzy clustering procedures the most mathematically strict are algorithms based on objective functions [Bezdek, 1981], that solve the problem of their optimization under various a priori assumptions. The most common is a probabilistic approach based on the minimization of the goal function

$$E\left(u_j, c_j\right) = \sum_{k=1}^{N}\sum_{j=1}^{m} u_j^{\beta}(k) x\left(k\right) - c_j^{\;2} \tag{1}$$

under constraints

$$\sum_{j=1}^{m} u_j(k) = 1, \tag{2}$$

$$0 \le \sum_{k=1}^{N} u_j\left(k\right) \le N, \tag{3}$$

where $u_j\left(k\right) \in [0,1]$ is the level of membership of vector $x(k)$ to k -th class $c_j -$, centroid (prototype) of j -th cluster, $\beta -$ a non-negative parameter of fuzzification (fuzzifier) defining boundaries blur between clusters, $k = 1, 2, \ldots, N$. The result of clustering is $\left(N \times m\right) -$ matrix $U = \left\{u_j(k)\right\}$, called the matrix of the fuzzy partitioning.

Let’s note that because of constraints (2), elements of the matrix U can be considered as the probabilities of hypotheses of data vectors membership to defined clusters, because of what the procedures generated by minimizing (1) are called probabilistic algorithms of fuzzy clustering. The number of clusters m is given a priori and cannot be changed during the computation.

Introducing the Lagrange function

$$L\left(u_j\left(k\right), c_j, \lambda(k)\right) = \sum_{k=1}^{N}\sum_{j=1}^{m} u_j^{\beta}\left(k\right) x\left(k\right) - c_j^{\;2} + \sum_{k=1}^{N} \lambda\left(k\right)\left(\sum_{j=1}^{m} u_j\left(k\right) - 1\right) \tag{4}$$

(there $\lambda\left(k\right) -$ undetermined Lagrange multiplier) and solving Karush-Kuhn-Tucker system of equations, we can easily obtain the desired solution in the form

$$\begin{cases} u_j\left(k\right) = \dfrac{\left(x\left(k\right) - c_j^{\;2}\right)^{\frac{1}{1-\beta}}}{\sum_{l=1}^{m}\left(x\left(k\right) - c_l^{\;2}\right)^{\frac{1}{1-\beta}}}, \\ c_j = \dfrac{\sum_{k=1}^{N} u_j^{\beta}\left(k\right) x(k)}{\sum_{k=1}^{N} u_j^{\beta}\left(k\right)}, \\ \lambda\left(k\right) = -\left(\left(\sum_{l=1}^{m} \beta x\left(k\right) - c_l^{\;2}\right)^{\frac{1}{1-\beta}}\right)^{1-\beta}, \end{cases} \tag{5}$$

which at $\beta = 2$ coincides with Bezdek’s Fuzzy C-means algorithm (FCM) [Bezdek, 1981], and at $\beta \to 0$ is close to results obtained with the help of popular algorithm of ordinary (hard) the k-means clusterization (HCM) [McQueen, 1965].

Algorithm (5) cannot be in the fullest sense called fuzzy, because it contains crisp value of fuzzifier $1 < \beta < \infty$, usually chosen from empirical considerations and essentially impact on the results. In this case variation of β from 1 to ∞ corresponds to transition from absolutely crisp boundaries ($\beta \to 1$) to total their blurring ($\beta \to \infty$) when all the observations belong to all clusters with the same level of membership.

Due to this Klawon and Höppner [Klawonn, Höppner, 2003] have proposed to use following expression for solving the problem of fuzzy clustering instead of the criterion (1) with a crisp fuzzifier:

$$E\left(u_j, c_j\right) = \sum_{k=1}^{N}\sum_{j=1}^{m}(\alpha u_j^2(k) + (1-\alpha)u_j(k))x\left(k\right) - c_j^{\ 2} \tag{6}$$

with restrictions (2), (3), where $0 < \alpha \le 1$ is a tunable parameter that determines character of solution obtained.

Introducing the Lagrange function

$$L\left(u_j\left(k\right), c_j, \lambda\left(k\right)\right) = \sum_{k=1}^{N}\sum_{j=1}^{m}\left(\alpha u_j^2\left(k\right) + \left(1-\alpha\right)u_j\left(k\right)\right)x\left(k\right) - c_j^{\ 2} + \sum_{k=1}^{N}\lambda\left(k\right)\left(\sum_{j=1}^{m}u_j\left(k\right) - 1\right) \tag{7}$$

and solving the system of Karush-Kuhn-Tucker equations

$$\begin{cases} \dfrac{\partial l\left(u_j\left(k\right), c_j, \lambda\left(k\right)\right)}{\partial u_j\left(k\right)} = \left(2\alpha u_j\left(k\right) + 1 - \alpha\right)x\left(k\right) - c_j^{\ 2} + \lambda\left(k\right) = 0, \\ \nabla_{c_j} L\left(u_j\left(k\right), c_j, \lambda\left(k\right)\right) = -\sum_{k=1}^{N} 2\left(\alpha u_j^2\left(k\right) + \left(1-\alpha\right)u_j\left(k\right)\right)\left(x\left(k\right) - c_j\right) = \vec{0}, \\ \dfrac{\partial L\left(u_j\left(k\right), c_j, \lambda\left(k\right)\right)}{\partial \lambda\left(k\right)} = \sum_{j=1}^{m} u_j\left(k\right) - 1 = 0, \end{cases} \tag{8}$$

we obtain the solution

$$\begin{cases} u_j\left(k\right) = -\dfrac{1-\alpha}{2\alpha} + \dfrac{1 + m\dfrac{1-\alpha}{2\alpha}}{\sum_{l=1}^{m}\dfrac{x\left(k\right) - c_j^{\ 2}}{x\left(k\right) - c_l^{\ 2}}}, \\ c_j = \dfrac{\sum_{k=1}^{N}(\alpha u_j^2\left(k\right) + (1-\alpha)u_j(k))x(k)}{\sum_{k=1}^{N}(\alpha u_j^2\left(k\right) + (1-\alpha)u_j(k))}, \\ \lambda\left(k\right) = -\dfrac{1 + m\dfrac{1-\alpha}{2\alpha}}{\sum_{l=1}^{m}\left(2\alpha x\left(k\right) - c_l^{\ 2}\right)^{-1}}. \end{cases} \tag{9}$$

It is easy to see that when $\alpha = 1$ we obtain the Bezdek's FCM algorithm

$$\begin{cases} u_j^{FCM}(k) = \dfrac{\|x(k)-c_j\|^{-2}}{\sum_{l=1}^{m}\|x(k)-c_l\|^{-2}}, \\ c_j^{FCM} = \dfrac{\sum_{k=1}^{N}(u_j^{\beta}(k))^2 x(k)}{\sum_{k=1}^{N}(u_j^{\beta}(k))^2}. \end{cases} \tag{10}$$

Taking that into account the first relation of (9) can be rewritten as

$$u_j(k) = -\frac{1-\alpha}{2\alpha} + \left(1 + m\frac{1-\alpha}{2\alpha}\right)u_j^{FCM}(k). \tag{11}$$

It is interesting to note that when $\alpha = \frac{1}{3}$ we obtain a compact expression

$$u_j(k) = (1+m)u_j^{FCM}(k) - 1. \tag{12}$$

Using all of above procedures assumes that the sample to be clustering containing N observations is given beforehand and cannot be changed during operation. In [Park, Dagher, 1984; Chung, 1994; Бодянский, Горшков, Кокшенев, Колодяжный, 2002; Bodyanskiy, Kolodyaznhiy, Stephan, 2002; Bodyanskiy, 2005] a group of recurrent adaptive procedures that allow to solve the problem of clustering in on-line mode was introduced.

Applying to (7) the Arrow-Hurwicz-Uzava nonlinear programming procedure, we come to an adaptive algorithm of fuzzy clustering according to the criterion (6):

$$\begin{cases} u_j(k+1) = -\dfrac{1-\alpha}{2\alpha} + \dfrac{1 + m\dfrac{1-\alpha}{2\alpha}}{\sum_{l=1}^{m}\dfrac{\|x(k+1)-c_j(k)\|^2}{\|x(k+1)-c_l(k)\|^2}}, \\ c_j(k+1) = c_j(k) + \eta(k)\left(\alpha u_j^2(k+1) + (1-\alpha)u_j(k+1)\right)\left(x(k+1)-c_j(k)\right), \end{cases} \tag{13}$$

where $\eta(k)$ – the learning parameter.

Possibilistic fuzzy clustering

The main disadvantages of the probabilistic approach associated with the constraint (2) requiring that the sum of memberships of each observation to all clusters would equal to one and the number of clusters m would set a priori. These disadvantages do not have methods of possibilistic fuzzy clustering, which were originally proposed by Krishnapuram and Keller [Krishnapuram, Keller, 1993].

In possibilistic clustering algorithms, the objective function has the form:

$$E\left(u_j, c_j\right) = \sum_{k=1}^{N}\sum_{j=1}^{m} u_j^{\beta}\left(k\right) x\left(k\right) - c_j^{\,2} + \sum_{j=1}^{m}\mu_j \sum_{k=1}^{N}\left(1-u_j(k)\right)^{\beta}, \tag{14}$$

where the scalar parameter $\mu_j > 0$ determines the distance at which level of membership takes the value 0.5, i.e. if

$$x\left(k\right) - c_j^{\,2} = \mu_j, \tag{15}$$

then $u_j\left(k\right) = 0,5$.

Direct minimization of (14) on $u_j\left(k\right)$ and c_j gives the known solution

$$\begin{cases} u_j\left(k\right) = \left(1 + \left(\dfrac{x\left(k\right) - c_j^{\,2}}{\mu_j}\right)^{\frac{1}{\beta-1}}\right)^{-1}, \\ c_j = \dfrac{\sum_{k=1}^{N} u_j^{\beta}\left(k\right) x(k)}{\sum_{k=1}^{N} u_j^{\beta}\left(k\right)}, \\ \mu_j\left(k\right) = \dfrac{\sum_{k=1}^{N} u_j^{\beta}(k) x\left(k\right) - c_j^{\,2}}{\sum_{k=1}^{N} u_j^{\beta}\left(k\right)}. \end{cases} \tag{16}$$

Using instead of (14) criterion

$$E\left(u_j, c_j\right) = \sum_{k=1}^{N}\sum_{j=1}^{m}\left(\alpha u_j^2\left(k\right) + \left(1-\alpha\right) u_j\left(k\right)\right) x\left(k\right) - c_j^{\,2} + \\ + \sum_{j=1}^{m}\mu_j \sum_{k=1}^{N}\left(\alpha(1-u_j\left(k\right))\right)^2 + (1-\alpha)(1-u_j\left(k\right)) \tag{17}$$

leads to a possibilistic procedure with variable fuzzifier

$$\begin{cases} u_j(k) = \dfrac{\mu_j(1+\alpha) - (1-\alpha)\|x(k) - c_j\|^2}{2\alpha\left(\|x(k) - c_j\|^2 + \mu_j\right)}, \\ c_j = \dfrac{\sum_{k=1}^{N}(\alpha u_j^2(k) + (1-\alpha)u_j(k))x(k)}{\sum_{k=1}^{N}(\alpha u_j^2(k) + (1-\alpha)u_j(k))}, \\ \mu_j = \dfrac{\sum_{k=1}^{N}\left(\alpha u_j^2(k) + (1-\alpha)u_j(k)\right)\|x(k) - c_j\|^2}{\sum_{k=1}^{N}\left(\alpha u_j^2(k) + (1-\alpha)u_j(k)\right)}. \end{cases} \quad (18)$$

In adaptive variant with processing sequentially receiving data we obtain at the recurrent procedure

$$\begin{cases} u_j(k+1) = \dfrac{\mu_j(k)(1+\alpha) - (1-\alpha)\|x(k+1) - c_j(k)\|^2}{2\alpha\left(\|x(k+1) - c_j(k)\|^2 + \mu_j(k)\right)}, \\ c_j(k+1) = c_j(k) + \eta(k)\left(\alpha u_j^2(k+1) + (1-\alpha)u_j(k+1)\right)\left(x(k+1) - c_j(k)\right), \\ \mu_j(k+1) = \dfrac{\sum_{p=1}^{k+1}\left(\alpha u_j^2(p) + (1-\alpha)u_j(p)\right)\|x(p) - c_j(k+1)\|^2}{\sum_{p=1}^{k+1}\left(\alpha u_j^2(p) + (1-\alpha)u_j(p)\right)} \end{cases} \quad (19)$$

Let’s note that in (13) and (19) algorithms for tuning the centroids of clusters are identical, and the difference consists in computation of membership levels to concrete cluster. It is also important that in contrast to probabilistic algorithms, possibilistic procedures allow during data processing to detect the appearance of new clusters. For example, if the level of membership of observation $x(k+1)$ for all clusters is below some preset threshold, we can speak about appearance of a $(m+1)$-th cluster with the initial coordinates of the centroid $c_{m+1} = x(k+1)$.

Adaptive neuro-fuzzy Kohonen network

It is easy to see that the second relations in the systems of equations (13), (19), being rewritten in the form

$$c_j(k+1) = c_j(k) + \eta(k)\varphi_j(k+1)\left(x(k+1) - c_j(k)\right), \quad (20)$$

represent nothing else but setting rules of self-organizing Kohonen neural network based on the principle of "winner takes more" [Kohonen, 1995], where $\varphi_j(k+1)$ is a neighborhood function. For these reasons, it is convenient to solve the considered problem of fuzzy clusterization on a basis of adaptive neuro-fuzzy Kohonen network, which architecture is shown in Fig. 1 and is a generalization of the architectures introduced in [Bodyanskiy, Gorshkov, Kolodyaznhiy, Stephan, 2005; Gorshkov, Kolodyaznhiy, Bodyanskiy, 2009].

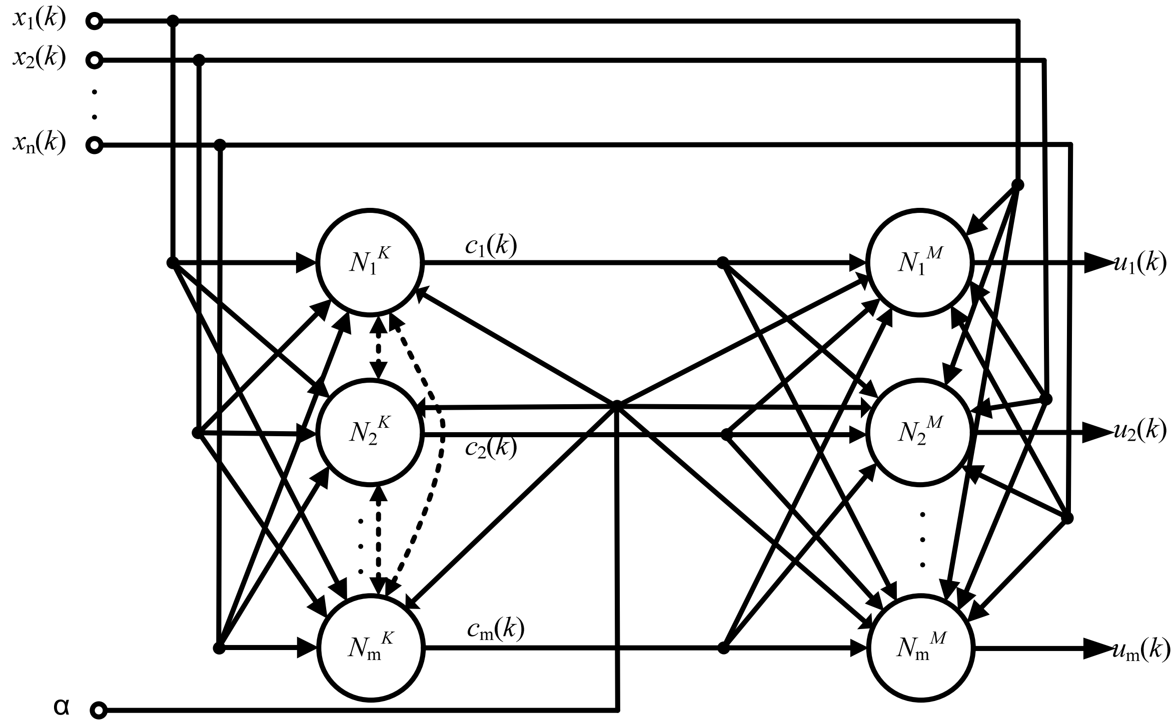

Figure 1 – Adaptive neuro-fuzzy Kohonen network with a variable fuzzifier

This network contains two layers: the Kohonen layer, which defines the prototypes (centroids) of clusters, and the layer of membership computing. Input vectors-pattern $x(k)=(x_1(k),\ x_2(k),\dots,x_n(k))^T$ from reception (zero) layer are sequentially fed to the neurons of Kohonen layer N_j^K, that are tuned by the rule (20). Synaptic weights $c_{ji}(k),\ j=1,2,\dots,m; i=1,2,\dots,n$ define centroids of m overlapping clusters $c_j(k)=(c_{j1}(k),\dots,c_{ji}(k),\dots,c_{jn}(k))^T$. In the output layer formed by neurons N_j^M levels of membership $u(k)=(u_1(k),\ u_2(k),\dots,u_n(k))^T$ of the presented vector $x(k)$ to j-th cluster according to the first relations of systems (13), (19) are calculated.

At lateral connections of Kohonen layer (shown dashed) processes of competition and cooperation are implemented, that are inherent to process of the Kohonen network training. To an additional input of the zero layer a customizable value of parameter α is fed.

Conclusion

The problem of fuzzy clustering of multidimensional observations with variable fuzzifier is considered and group of adaptive algorithms which enable to process data in real time as it received is proposed. Introduced algorithm is characterized by numerical simplicity and offers more flexibility when working in conditions of a priori uncertainty about the nature of data distribution in clusters. Proposed algorithms are used as learning rules of self-organizing neuro-fuzzy Kohonen network.

Acknowledgements

The paper is published with financial support by the project ITHEA XXI of the Institute of Information Theories and Applications FOI ITHEA (www.ithea.org) and the Association of Developers and Users of Intelligent Systems ADUIS Ukraine (www.aduis.com.ua).

Bibliography

[Tukey, 1977] Tukey J. W. Exploratory Data Analysis. Reading, MA: Addison-Wesley Publ. Company, Inc., 1977.

[Höppner, Klawonn, Kruse, Runkler, 1999] Höppner F., Klawonn F., Kruse R., Runkler T. Fuzzy Clustering Analysis: Methods for Classification, Data Analysis and Image Recognition. Chichester: John Willey & Sons., 1999,

[Bezdek, 1981] Bezdek J.C. Pattern Recognition with Fuzzy Objective Function Algorithms, N.Y.:Plenum Press., 1981.

[Gath, Geva, 1989] Gath I., Geva A.B. Unsupervised optimal fuzzy clustering In: Pattern Analysis and Machine Intelligence., 1989., 2., 7., P. 773-787

[McQueen, 1965] McQueen J. On convergence of k-means and partitions with minimum average variance In: Ann. Math. Statist., 1965., 36., P. 1084.

[Klawonn, Höppner, 2003] Klawonn F., Höppner F. What is fuzzy about fuzzy clustering? Understanding and improving the concept of the fuzzifier. In: Lecture Notes in Computer Science., Springer Berlin-Heidelberg., 2003., Vol. 2811., P. 254-264.

[Park, Dagher, 1984] Park D. C., Dagher I. Gradient based fuzzy c-means (GBFCM) algorithm In: Proc. IEEE Int. Conf. on Neural Networks., 1984., P. 1626-1631.

[Chung, 1994] Chung F.L., Lee T. Fuzzy competitive learning In: Neural Networks., 1994., 7., №3., P. 539-552.

[Бодянский, Горшков, Кокшенев, Колодяжный, 2002] Бодянский Е.В., Горшков Е.В., Кокшенев И.В, Колодяжный В. В. Об адаптивном алгоритме нечёткой кластеризации данных In: Адаптивні системи автоматичного управління., Вип.5(25)., Дніпропетровськ: Системні технології., 2002., С. 108-117.

[Bodyanskiy, Kolodyaznhiy, Stephan, 2002] Bodyanskiy Ye., Kolodyaznhiy V., Stephan A. Recursive fuzzy clustering algorithms In: Proc. 10th East West Fuzzy Colloqium., Zittau, Germany., 2002., P. 276-283.

[Bodyanskiy, 2005] Bodyanskiy Ye. Computational intelligence techniques for data analysis In: Lecture Notes in Informatics., Bonn, Germany., 2005., Vol. P-72., P. 15-36.

[Krishnapuram, Keller, 1993] Krishnapuram R., Keller J. M. A possibilistic approach to clustering In: Fuzzy Systems., 1993., 1., №2., P. 98-110.

[Kohonen, 1995] Kohonen T. Self-Organizing Maps / Kohonen T., Berlin: Springer-Verlag., 1995.

[Bodyanskiy, Gorshkov, Kolodyaznhiy, Stephan, 2005] Bodyanskiy Ye. Combined learning algorithm for a self-orginizing map with fuzzy inference In: B. Reusch (Ed) "Computational intelligence: theory and applications"., Advanced in Soft Computing, Vol. 3: Berlin-Heidelberg: Springer-Verlag, 2005., P. 641-650.

[Gorshkov, Kolodyaznhiy, Bodyanskiy, 2009] Gorshkov Ye. New recursive learning algorithms for fuzzy Kohonen clustering network. In: Proc. 17th Int. Workshop on Nonlinear Dynamics of Electronic Systems., Rapperswil, Switzerland, 2009., P. 58-61.

Authors' Information

Yevgeniy Bodyanskiy – Professor, Dr.-Ing. habil., Scientific Head of Control Systems Research Laboratory, Kharkiv National University of Radio Electronics; 14 Lenin Ave., Office 511, 61166 Kharkiv, Ukraine; e-mail: bodya@kture.kharkov.ua

Major Field of Scientific Research: hybrid systems of computational intelligence.

Bogdan Kolchygin – Ph.D. student, Junior Researcher of Control Systems Research Laboratory, Kharkiv National University of Radio Electronics; 14 Lenin Ave., Office 511, 61166 Kharkiv, Ukraine; e-mail: quasimail@gmail.com

Major Field of Scientific Research: adaptive fuzzy clustering.

Iryna Pliss – Ph.D., Leading Researcher of Control Systems Research Laboratory, Kharkiv National University of Radio Electronics; 14 Lenin Ave., Office 511, 61166 Kharkiv, Ukraine; e-mail: pliss@kture.kharkov.ua

Major Field of Scientific Research: hybrid systems of computational intelligence.