Hybrid Clustering-Classification Neural Network in the Medical Diagnostics of the Reactive Arthritis

Yevgeniy Bodyanskiy, Olena Vynokurova
Control Systems Research Laboratory, Kharkiv National University of Radio Electronics, Kharkiv, 61166, Ukraine
E-mail: {yevgeniy.bodyanskiy, olena.vynokurova}@nure.ua

Volodymyr Savvo, Tatiana Tverdokhlib
Pediatrics Department, Kharkiv Medical Academia of Post-Graduate Education, Kharkiv, 61176, Ukraine
E-mail: savvovm50@gmail.com, tanyamar82@gmail.com

Pavlo Mulesa
Cybernetics and Applied Mathematics Department, Uzhhorod National University, Uzhhorod, 88000, Ukraine
E-mail: ppmulesa@gmail.com

Abstract—In the paper, the hybrid clustering-classification neural network is proposed. This network allows to increase a quality of information processing under the condition of overlapping classes due to the rational choice of learning rate parameter and introducing special procedure of fuzzy reasoning in the clustering-classification process, which occurs both with external learning signal ("supervised"), and without one ("unsupervised"). As similarity measure neighborhood function or membership one, cosine structures are used, which allow to provide a high flexibility due to self-learning-learning process and to provide some new useful properties. Many realized experiments have confirmed the efficiency of proposed hybrid clustering-classification neural network; also, this network was used for solving diagnostics task of reactive arthritis.

Index Terms—Hybrid clustering-classification neural network, supervised/unsupervised learning, overlapping classes, diagnostics, reactive arthritis.

I. INTRODUCTION

The current state of technological development connects with the development of computer tools. These tools depend on the mathematical methods and the practical algorithms. The development of computer tools is activator for evolution of existing scientific areas and appearance of new scientific direction, for example, Data Science, Big Data, Computational Intelligence, Data Stream Mining etc. Modern capabilities of computing environments allow implementation of algorithmically sufficient complex methods, which are basis for Data Mining.

The hybrid neural networks are one of the most wide spread approaches for solving of Data Mining tasks. Self-organizing maps (SOM) and neural networks of learning vector quantization (LVQ) have seen extensive use for solving different problems in Data Mining domain (clustering, classification, fault detection and compression of information etc.). This type of neural networks was proposed by T. Kohonen [1, 2] and represents, in fact, a single-layer feedforward architecture, which provides an operator for mapping of input space into the output space. Operation-wise SOM and LVQ are quite similar to each neuron is fed input signal (sample) producing output, which is used during competition stage to determine winning neuron – usually the one with maximum output signal value. Vector of synaptic weights for winning neuron is the one closest to the input sample in terms of the metric chosen (which is Euclidian metric in most cases). Next is neurons adjustment phase. Synaptic weights of the winning neuron gets moved closer to input sample. Alternatively, a subset of neurons (rather than a single one) can be adjusted – those determined to be "reasonably close" to the input sample are updated. Resulting network output, however, is determined exclusively by winning neuron (this principle is usually referred to as "Winner-Takes-All" (WTA)). It is this principle (WTA) which negatively affects accuracy in case when there are overlapping clusters in underlying data.

II. RELATED WORK

Taking into account the above mentioned properties of SOM and LVQ networks, it makes sense to introduce fuzzy classification capabilities on top of them, while preserving online operation.

In [8, 9] fuzzy self-organizing map was proposed, in which conventional neurons are replaced by fuzzy rules. This neural network shows enough high efficiency, but its learning properties were significantly lost especially in on-line mode.

In [5, 10, 11] fuzzy clustering Kohonen network and fuzzy linear vector quantization network are described. In fact, such networks are neural representation of fuzzy c-means (FCM) [3], which is far enough from SOM and LVQ mathematical tool and designed for operation in batch mode.

III. PROBLEM STATEMENT

Let us consider single-layer neural network with lateral connections containing receptors and neurons in the Kohonen layer with each neuron being characterized by a vector of its synaptic weights. During learning stage input vector is fed to the inputs of all neurons (usually adaptive linear associators) (here - either the number of observation in a table "object-properties", or current discrete time for on-line processing mode) and neurons produce the scalar signals on their outputs

$$y_j(k) = w_j^T(k)x(k), \ \ j = 1, 2, ..., m. \tag{1}$$

Note that neuron's output depends on current values of synaptic weights vector, assuming iterative learning algorithm.

Each input vector can activate either single neuron (w_j) or a set of neighboring neurons – this also depends on learning algorithm chosen.

Self-organization procedure is based on the competitive learning approach (self-learning) and begins with the initialization of synaptic weights. Selecting initial values for weights from a uniform random distribution over input space has become a common practice. In addition, weights are usually normalized to reside on unit hypersphere:

$$\left\| w_j(k) \right\|^2 = w_j^T(k)w_j(k) = 1. \tag{2}$$

Self-organizing usually contains three stages [2]: competition, cooperation, and synaptic adaptation. This paper introduces additional stage to this process – fuzzy inference, which allows an online learning algorithm to classify data samples belonging to several classes simultaneously.

IV. LEARNING ALGORITHM FOR SOM

The competition process is started to analysis of current pattern $x(k)$, which is fed to all neurons of Kohonen's layer from receptive (zero) layer. For each neuron the distance between input sample and a vector of synaptic weights is computed:

$$D(w_j(k), x(k)) = \left\| x(k) - w_j(k) \right\|. \tag{3}$$

In case when all inputs and synaptic weights are normalized according to (2), i.e.

$$\left\| w_j(k) \right\| = \left\| x(k) \right\| = 1, \tag{4}$$

and Euclidian metric is used, the proximity measure between the vectors $w_j(k)$ and $x(k)$ can be written in the following way:

$$w_j^T(k)x(k) = y_j(k) = \cos(w_j(k), x(k)) = \cos\theta_j(k). \tag{5}$$

So the expression (3) takes the form

$$D(w_j(k), x(k)) = \sqrt{2(1 - y_j(k))}. \tag{6}$$

Using relation (5), we can replace metric (3) with a simpler construction, usually referred to as a measure of similarity [12]

$$\psi(w_j(k), x(k)) = \max\left\{0, \cos\theta_j(k)\right\}, \tag{7}$$

which satisfies softer conditions compared to classic measure requirements:

$$\begin{aligned} &\psi(w_j(k), x(k)) \geq 0, \\ &\psi(w_j(k), w_j(k)) = 1, \\ &\psi(w_j(k), x(k)) = \psi\left\{x(k), w_j(k)\right\}. \end{aligned} \tag{8}$$

Next we look for a winning neuron that has a biggest value of similarity to the input vector in the sense (6) or (7), i. e.

$$D(w_j^*(k), x(k)) = \min_l D(w_l(k), x(k)) \tag{9}$$

or

$$\psi(w_j^*(k), x(k)) = \max_l \psi(w_l(k), x(k)). \tag{10}$$

After that, we tune values of synaptic weights using WTA self-learning rule in form

$$w_j(k+1) = \begin{cases} w_j^*(k) + \eta(k)(x(k) - w_j^*(k)), \\ \qquad \text{if } j - \text{th neuron won in} \\ \qquad\quad \text{the competition,} \\ w_j(k) \ \ \text{otherwise.} \end{cases} \tag{11}$$

In a nutshell, expression (11) means that synaptic weights vector of winner is moved closer to the input pattern $x(k)$ by a value which is proportional to learning rate $0 < \eta(k) < 1$.

Choice of learning rate value $\eta(k)$ defines the convergence rate of self-learning process, and is usually based on empirical reasons, at that the parameter must monotonically decrease in on time.

It is easy to see, that first relation of the rule (11)

minimizes the criterion

$$E_j^k = \sum_{\tau=1}^{k_j} \left\| x(\tau) - w_j \right\|^2 \qquad (12)$$

(here k_j is a number of data in the sampling with dimensions k, when j-th neuron was winner), notably in practice as synaptic weights estimation conventional arithmetic mean is used:

$$w_j(k) = \frac{1}{k_j} \sum_{\tau=1}^{k_j} x(\tau). \qquad (13)$$

Therefore, in fact, self-learning rule (3) is stochastic approximation procedure [13], and learning rate parameter $\eta(k)$ must be selected according to the conditions of A. Dvoretzky [14]. It is known that using the following value

$$\eta(k) = \frac{1}{k_j} \qquad (14)$$

leads to slowdown of the learning process.

Requirement of monotonic decreasing of the parameter $\eta(k)$ leads to expression in form

$$\eta(k) = r^{-1}(k),$$

$$r(k) = \alpha r(k-1) + \left\| x(k) \right\|^2, \ 0 \leq \alpha \leq 1. \qquad (15)$$

When taking into account normalization to unit hypersphere (4) we have:

$$r(k) = \alpha r(k-1) + 1, \qquad (16)$$

which with $\alpha = 1$ gives a well-known expression, that was introduced in [15].

Changing the forgetting parameter α provides enough variance for the learning rate to both satisfy Dvoretzky condition and adjust for characteristics of different data sets:

$$\frac{1}{k_j} \leq \eta(k) \leq 1. \qquad (17)$$

Note that adjusting synaptic weights with (13) in general breaks normalization (4). In order to maintain it, we should modify our learning algorithm:

$$w_j(k+1) = \begin{cases} \dfrac{w_j^*(k) + \eta(k)(x(k) - w_j^*(k))}{\left\| w_j^*(k) + \eta(k)(x(k) - w_j^*(k)) \right\|}, \\ \text{if } j-th \text{ neuron won in the} \\ \text{competition,} \\ \eta(k) = r^{-1}(k), \\ r(k) = \alpha r(k-1) + 1, \ 0 \leq \alpha \leq 1, \\ w_j(k), \text{ otherwise.} \end{cases} \qquad (18)$$

In order to better adjust to input data, a so-called "cooperation stage" is frequently introduced to SOM learning process. During this stage, a winning neuron defines a local domain of topological neighborhood, where weights are adjusted for a set of neurons rather than only for a winning one. Those neurons closer to the winner receive a bigger adjustment.

This topological domain is defined by neighborhood function $\varphi(j, p)$, which depends on distance $D(w_j^*(k), w_p(k))$ between winner $w_j^*(k)$ and any of Kohonen's layer neurons $w_p(k)$, $p = 1, 2, \ldots, m$ and some parameter σ, which sets its width.

Usually $\varphi(j, p)$ is the bell-shaped function, symmetrical with respect to the maximum in point $w_j^*(k)$ ($\varphi(j, j) = 1$) and decreasing when distance $D(w_j^*(k), w_p(k))$ increases. Gaussian function is commonly used here:

$$\varphi(j, p) = \exp\left(-\frac{\left\| w_j^*(k) - w_p(k) \right\|^2}{2\sigma^2} \right). \qquad (19)$$

Adding neighborhood function results in the following learning algorithm:

$$w_p(k+1) = w_p(k) + \eta(k)\varphi(j, p)(x(k) - w_p(k)), \qquad (20)$$

which minimizes criterion

$$E_p^k = \sum_{\tau=1}^{k} \varphi(j, p) \left\| x(\tau) - w_p \right\|^2 \qquad (21)$$

This criterion is commonly referred to as "Winner Takes More" (WTM).

The necessity to maintain normalization to unit hypersphere (4) leads to the expression in form

$$\begin{cases} w_p(k+1) = \dfrac{w_p(k) + \eta(k)\varphi(j, p)(x(p) - w_p(k))}{\left\| w_p(k) + \eta(k)\varphi(j, p)(x(p) - w_p(k)) \right\|}, \\ p = 1, 2, \ldots, m, \\ \\ \eta(k) = r^{-1}(k), \ r(k) = \alpha r(k-1) + 1, 0 \leq \alpha \leq 1. \end{cases} \qquad (22)$$

It is possible to skip the entire competition by tuning synaptic weights of network depending on their similarity with the current vector-pattern $x(k)$. In this case instead of conventional neighborhood function $\varphi(j, p)$ depending on winner $w_j^*(k)$, we can use similarity measure (7), that depends on $x(k)$.

In this case instead of (20) we can use the rule in form:

$$
\begin{aligned}
w_p(k+1) &= w_p(k) + \\
&\quad + \eta(k)\psi(w_p(k), x(k))(x(k) - w_p(k)) = \quad (23) \\
&= w_p(k) + \eta(k)\max\{0, y_p(k)\}(x(k) - w_p(k)),
\end{aligned}
$$

which reminds «INSTAR» learning rule of S. Grossberg [16].

For maintaining (4) the rule (23) has to be rewritten in the form

$$
\begin{cases}
w_p(k+1) = \dfrac{w_p(k) + \eta(k)\psi(w_p(k), x(k))(x(k) - w_p(k))}{\left\| w_p(k) + \eta(k)\psi(w_p(k), x(k))(x(k) - w_p(k)) \right\|}, \\
\eta(k) = r^{-1}(k),\ r(k) = \alpha r(k-1) + 1,\ 0 \le \alpha \le 1.
\end{cases} \quad (24)
$$

In many real world problems clusters can overlap as shown on Fig. 1, where * denotes patterns, belonging to j-th clusters, and o − p-th ones. In this case vector $x(k)$ belongs j-th cluster with proportional membership level $\cos\theta_j(k)$, and with proportional level $\cos\theta_p(k)$ - to p-th one. Synaptic weights concentrated in crosshatched region, don't have relationship to the pattern $x(k)$ according to (7).

Using similarity measure, that is shown on fig.2, we can introduce the membership estimate of pattern $x(k)$ to j-th class in form:

$$
0 \le \mu_{w_j(k)}(x(k)) = \frac{\psi(w_j(k), x(k))}{\displaystyle\sum_{l=1}^{m} \psi(w_l(k), x(k))} \le 1. \quad (25)
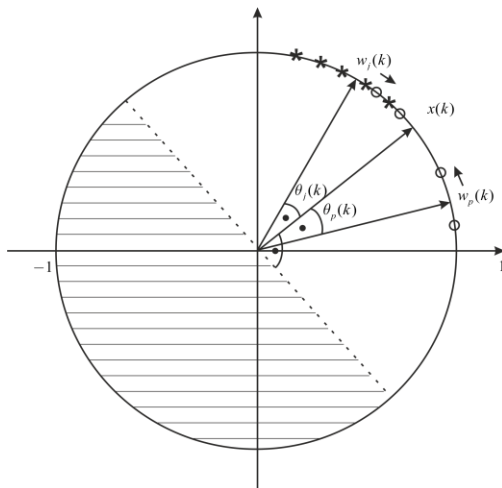$$

Fig.1. The overlapping clusters

V. LEARNING ALGORITHM FOR LVQ

Learning vector quantization neural networks in contrast to self-learning SOM adjust their parameters based on external learning (reference) signal, which fixes the membership of each pattern $x(k)$ to a particular class.

The main idea of LVQ neural network is to build a compact representation of large data set based on a restricted set of prototype samples (centroids) $w_j(k)$, $j = 1, 2, \ldots, m$, that provide sufficiently accurate approximation of the original space X.

For each input vector $x(k)$ which is normalized according to (4), we look for a winning neuron $w_j^*(k)$ that corresponds to a centroid of a certain class. In other words, winner is defined as a neuron with minimal distance to the input vector (9) or, which is the same, with maximal similarity measure (10).

Since the learning is supervised, membership of the vector $x(k)$ to specific domain X_j of the space X is known, which allows considering two typical situations, which occur in the vector quantization:

- the input vector $x(k)$ and neuron-winner $w_j^*(k)$ belong to the same cell of Voronoy [2];
- the input vector $x(k)$ and neuron-winner $w_j^*(k)$ belong to the different cells of Voronoy.

Than corresponding learning LVQ-rule can be written in form:

$$
w_j(k+1) = \begin{cases}
w_j^*(k) + \eta(k)(x(k) - w_j^*(k)), \\
\text{if } x(k) \text{ and } w_j^*(k) \text{ belong to the same cell,} \\[6pt]
w_j^*(k) - \eta(k)(x(k) - w_j^*(k)), \\
\text{if } x(k) \text{ and } w_j^*(k) \text{ belong to the} \\
\text{different cells,} \\[6pt]
w_j(k) \text{ for neurons,} \\
\text{which aren't won in instant } k.
\end{cases} \quad (26)
$$

The rule (26) has a clear physical interpretation: if the winning neuron and presented sample belong to the same class, than prototype $w_j^*(k)$ is moved to $x(k)$; otherwise prototype $w_j^*(k)$ is pushed away from $x(k)$, effectively increasing the distance $D(w_j^*(k), x(k))$, i.e. solving the maximization task based on criterion (12).

As for the choice of learning rate parameter $\eta(k)$, that common recommendations are the same that for SOM, i.e. the learning rate parameter must monotonically decrease in process controlled learning.

In [17] it was proved that LVQ tuning algorithm

converges in case of learning rate $\eta(k)$ satisfies Dvoretzky conditions. This, in turn, allows choosing $\eta(k)$ according to Goodwin-Ramadge-Caines algorithm [13], or in the previously defined form (16) with $\alpha = 1$, which is essentially the same. When $\alpha < 1$, the algorithm gets tracking properties and handles the case when class centroids are drifting.

For providing to fulfillment of the condition (4) it possible to introduce modification of rule (26) in the form (18):

$$w_j(k+1) = \begin{cases} \dfrac{w_j^*(k) + \eta(k)(x(k) - w_j^*(k))}{\left\| w_j^*(k) + \eta(k)(x(k) - w_j^*(k)) \right\|}, \\ \text{if } x(k) \text{ and } w_j^*(k) \text{ belong to the} \\ \text{same cell,} \\ \\ \dfrac{w_j^*(k) - \eta(k)(x(k) - w_j^*(k))}{\left\| w_j^*(k) - \eta(k)(x(k) - w_j^*(k)) \right\|}, \\ \text{if } x(k) \text{ and } w_j^*(k) \text{ belong to} \\ \text{the different cells,} \\ \\ w_j(k) \text{ for neurons, which} \\ \text{are not won in instant } k. \end{cases} \quad (27)$$

First and third expressions in formula (27) are completely identical to WTA – self-learning algorithm, which the second one represents a "push-back" scenario.

Let's consider a situation, shown in Fig. 2, when neuron $w_j^*(k)$ won the competition.
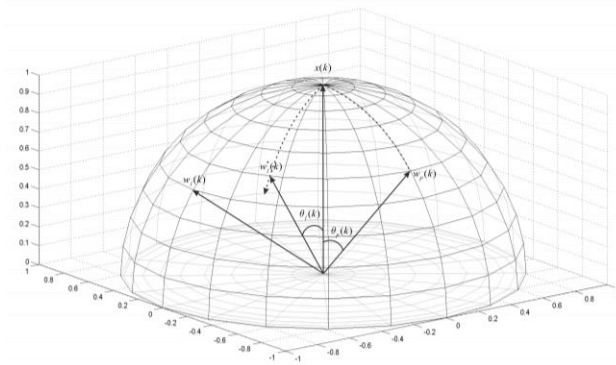

Fig.2. Learning of LVQ

At the same time current sample $x(k)$ belongs to a class with different centroid $w_p(k)$. Now we need to «push» vector $w_j^*(k)$ away so, that $x(k)$ was equidistant from $w_j^*(k)$ and from $w_p(k)$.

Applying elementary transformations, we get the following formulas:

$$w_j(k+1) = w_j^*(k) - \eta(k)(x(k) - w_j^*(k)) , \quad (28)$$

$$\begin{aligned} x^T(k)w_j(k+1) &= \\ = x^T(k)w_j^*(k) &- \eta(k)\|x(k)\|^2 - \eta(k)x^T(k)w_j^*(k) \end{aligned}, \quad (29)$$

$$\begin{aligned} \cos(w_j(k+1), x(k)) &= \\ = \cos(w_j^*(k), x(k)) &- \eta(k)(1 + \cos(w_j^*(k), x(k))) \end{aligned} \quad (30)$$

In order to satisfy

$$\cos(w_j(k+1), x(k)) = \cos(w_p(k), x(k)) \quad (31)$$

we need to set $\eta(k)$ in form

$$\begin{aligned} \eta(k) &= \frac{\cos(w_j^*(k), x(k)) - \cos(w_p(k), x(k))}{\cos(w_j^*(k), x(k)) + 1} = \\ &= \frac{x^T(k)w_j^*(k) - x^T(k)w_p(k)}{x^T(k)w_j^*(k) - x^T(k)x(k)}. \end{aligned} \quad (32)$$

After one step of the weights tuning the pattern $x(k)$ will belong equally to both $w_j(k+1)$ and $w_p(k) = w_p(k+1)$, i.e.

$$\mu_{w_j(k+1)}(x(k)) = \mu_{w_p(k+1)}(x(k)) = 0,5 . \quad (33)$$

In case when several classes are overlapping, we can use estimate (25) for computing membership levels.

VI. Hybrid Clustering-Classification Neural Network with Supervized/ Unsupervised Learing

A promising application of Kohonen neural network is adaptive pattern recognition, which implemented by the systems. The implementation consists of sequentially connected SOM and LVQ [2] layers. First part of network operates in self-learning mode, while the second one adds supervised component to the process.

From an input vector $x(k)$ which is fed to the input system, SOM network extracts a feature sample $y(k)$ in a space with sufficiently reduced dimensionality. This simplifies a problem in hand without significant loss of information.

On the second stage LVQ is trained to classify of incoming pattern $y(k)$ using supervised learning. A distinctive feature of proposed system is the connection of two identical architectures, that are tuned by different learning algorithms.

For many data mining problems, especially in healthcare domain, substantial shares of input samples lack clear class association. Indeed, a diagnosis might be either missing, be ambiguous or yet unknown.

In this case it is possible to tune the synaptic weights

with unified architecture with lateral connections using different learning methods. Each learning method is initialized according to the level of apriori information about $x(k)$ available.

Fig. 3 shows the architecture of hybrid clustering-classification neural network.
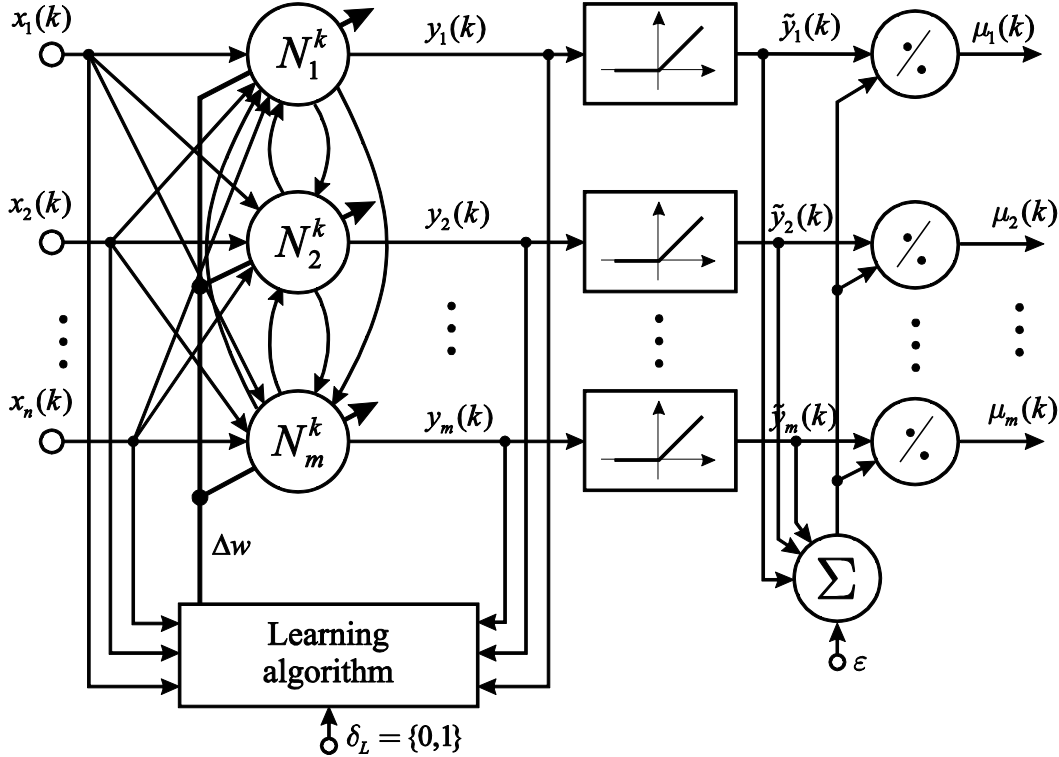

Fig.3. Architecture of hybrid clustering-classification neural network

A resulting learning algorithm for combined (SOM+LVQ) neural network can be presented in the following form [18]:

$$w_j(k+1) = \begin{cases} \dfrac{w_j^*(k)+\eta(k)(x(k)-w_j^*(k))}{\left\|w_j^*(k)+\eta(k)(x(k)-w_j^*(k))\right\|}, \\[2mm] \delta_L(k)\dfrac{w_j^*(k)-\eta(k)(x(k)-w_j^*(k))}{\left\|w_j^*(k)-\eta(k)(x(k)-w_j^*(k))\right\|}, \\[2mm] \delta_L = \begin{cases} 1, \text{ if the network works} \\ \quad \text{in supervised mode} \\ \quad \text{and sample x(k) does} \\ \quad \text{not belong to class j,} \\ 0, \text{ otherwise,} \end{cases} \\[2mm] w_j(k) \text{ for neurons, which did not} \\ \text{get activated by sample } x(k) \end{cases} \quad (34)$$

where $w_j^*(k)$ is winning neuron.

It is clear, that for $\delta_L = 0$, i.e. in self-learning mode, the first expression of formula (34) can be replaced by expression (24).

VII. RESULTS OF EXPERIMENT

Reactive arthritis (ReA) is the important medical-social problem of today's world because of high prevalence of ReA among people of different age groups. ReA occupies one of the top position among inflammatory disease of joints [19-21]. In different countries the frequency of ReA equal to from 8% to 41%.

According to some studies, ReA is the result of overproduction of proinflammatory cytokines or under ReA Th1-immune response is reduced in favor of Th2-immune response [22, 23]. There are scientific studies that prove the participation of cytokines in the pathogenesis of rheumatic diseases, the most common confirming the hypothesis of the pathogenesis of ReA, which is based on an imbalance of cytokines [24].

However, by now many researcher cannot establish reasons, which the infection process by the causative agent one in some individuals does not lead to medical problem, while others lead to the progress of infectious and inflammatory diseases in the acute or chronic form.

The study involved 150 children, including immunology research in 40 pediatric patients with ReA in acute form and after 9-12 months since the dawn of the disease.

Immunological research includes the study of measure of cellular, humoral, monocyte-phagocytic components of

immune system, the content of cytokines (IL-1β, IL-6). In cellular component of immune system the lymphocyte subpopulations was determined (CD3+, CD4+, CD8+, CD25, CD21). Determination of serum immunoglobulin of classes A, M, G is performed by spectrophotometry.

Monocyte-phagocytic component of immune system was evaluated based on the phagocytic and metabolic activity of leukocytes by determining the phagocytic number, spontaneous and stimulated NBT-test. For estimation of cytokine status in the examined patients the levels of IL-1β and IL-6 in serum is determined based on "sandwich"-method ELISA (enzyme linked immunosorbent assay) using monoclonal antibodies.

Therefore the data set is presented by the table "object-properties" consists of 40 objects and 12 properties. The hybrid clustering-classification neural network with learning algorithm (34) was used for solving of diagnosis task of ReA based on immunological indicators of pediatric patients. The initial values of synaptic weights were taken in a random way.

As the criterion of clustering-classification we used the percent of incorrect classified patterns in test sample.

The fig. 4 shows the results of clustering-classification of data, which describes immunological indicators of pediatric patients with ReA.

Table 1 shows the comparative analysis of clustering-classification results based on four approaches: proposed network with its learning algorithms, Kohonen self-organized map and LVQ-neural network with conventional learning algorithm and fuzzy C-means clustering algorithm.

Analysis of clustering-classification results shows the forming of two cluster, which are characterized two group of pediatric patients: group with prolonged, recrudescent and chronic form of ReA disease and group with recovering patients after proposed treatment.
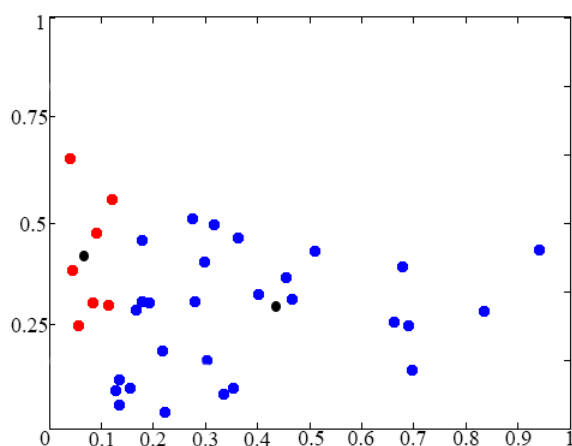

Fig.4. Results of clustering-classification data based on hybrid clustering-classification network (● – the cluster with prolonged, recrudescent and chronic form of ReA disease, ● – the cluster with recovering patients after treatment, ● – the prototypes of clusters)

The acute period of ReA in pediatric patients is characterized by the changes of immunological

homeostasis in form reducing of the relative content of CD8, CD25 in serum and increasing levels of CD21, IL-6, the phagocytic number and spontaneous NBT test.

Table 1. The comparative analysis of clustering-classification results

Neural network / Learning methods	Error of clustering-classification, train set	Error of clustering-classification, check set
Hybrid clustering-classification neural network / Proposed learning algorithm	3.1 %	4.5 %
Kohonen self-organized map / Conventional learning algorithm	6.9 %	8.0 %
LVQ neural network / Conventional learning algorithm	7.9 %	8.5 %
Fuzzy C-means clustering algorithm	6,8%	7,4%

Therefore, it is apparent that proposed approach provides best results of clustering-classification among considered approaches due to operation in learning/self-learning mode under missing of diagnosis data labeling or in the case when the diagnosis was not made because of assident diseases of patients.

Medical data mining result was diagnosis task solving for defining of hidden factors of reactive arthritis progression in the children, which allowed for appropriate treatment.

VIII. CONCLUSIONS

Methods of scientific direction that is developed within computer science and is known as computational intelligence have been widely used to solve various problems of data mining recently. The very first reason behind advantages of the computational intelligence is the fact that they perform data processing in biologically plausible way, leading to necessary flexibility, adaptivity, linguistic interpretability of the designed models.

Thus, in the paper the combined self-learning procedure for hybrid clustering-classification neural network is proposed. Such network allows data processing under the overlapping classes condition, when memberships of training data to specific classes can be unknown at all, and have both crisp and fuzzy nature. This procedure is based on using similarity measure, recurrent optimization and fuzzy inference and differs with high speed, possibility of operating in on-line mode and simplicity of computational realization. Medical data mining result was diagnosis task solving for defining of hidden factors of reactive arthritis progression in the children, which allowed for appropriate treatment.

REFERENCES

[1] T. Kohonen, *Self-Organizing Maps*. Berlin: Springer-Verlag, 1995. DOI: 10.1007/978-3-642-56927-2

[2] S. Haykin, *Neural Networks: A Comprehensive*

Foundation. Upper Saddle River, N.Y.: Prentice Hall, 1999.

[3] J.C. Bezdek, *Pattern Recognition with Fuzzy Objective Function Algorithms*. New York: Plenum Press, 1981. DOI: 10.1007/978-1-4757-0450-1

[4] F. Hoeppner, T. Klawon, and R. Kruse, *Fuzzy Clusteranalyse: Verfahren fuer die Belderkennung, Klassification und Datenanalyse.* – Braunschweig: Vieweg, Reihe Computational Intelligence, 1996. DOI: 10.1007/978-3-322-86836-7

[5] J.C. Bezdek, J. Keller, R. Krisnapuram, and N. Pal, *Fuzzy Models and Algorithms for Pattern Recognition and Image Processing*. Springer, 1999. DOI: 10.1007/b106267

[6] F. Hoeppner, F. Klawonn, R. Kruse, and T. Runkler, *Fuzzy Cluster Analysis: Methods for Classification, Data Analysis and Image Recognition*. Chichester: John Wiley & Sons, 1999.

[7] M. Sato-Ilic, and L. Jain, *Innovations in Fuzzy Clustering: Theory and Applications*. Berlin: Springer, 2006. DOI: 10.1007/3-540-34357-1

[8] P. Vuorimaa, "Use of the fuzzy self-organizing map in pattern recognition", in *Proc. 3-rd IEEE Int. Conf. Fuzzy Systems "FUZZ-IEEE'94"*, Orlando, USA, 1994, pp. 798-801. DOI: 10.1109/FUZZY.1994.343837

[9] P. Vuorimaa, Fuzzy self-organizing map. *Fuzzy Sets and Systems,* 1994, 66, pp. 223-231. DOI: 10.1016/0165-0114(94)90312-3

[10] E.C.-K. Tsao, J.C. Bezdek, and N.R. Pal, "Fuzzy Kohonen clustering networks," *Pattern recognition,* 1994, 27, pp. 757-764. DOI: 10.1016/0031-3203(94)90052-3

[11] R.D. Pascual-Marqui, A.D. Pascual-Montano, K. Kochi, and J.M. Caraso "Smoothly distributed fuzzy C-means: a new self-organizing map," *Pattern Recognition,* 2001, 34, pp. 2395-2402. DOI: 10.1016/S0031-3203(00)00167-9

[12] J.J. Sepkovski, "Quantified coefficients of association and measurement of similarity," *J. Int. Assoc. Math.*, 1974, 6 (2), pp. 135-152. DOI: 10.1007/BF02080152

[13] M.T. Wasan, *Stochastic Approximation*. Cambridge: The University Press, 2004. DOI: 10.1002/zamm.19700500635

[14] A. Dvoretzky, "Stochastic approximation revisited," *Advances in Applied Mathematics*, 1986, 7(2), pp. 220-227. DOI: 10.1016/0196-8858(86)90033-3

[15] G.C. Goodwin, P.J. Ramadge, and P.E. Caines, "A globally convergent adaptive predictor," *Automatica*, 1981, 17(1), pp. 135-140. DOI: 10.1016/0005-1098(81)90089-3

[16] S. Grossberg, "Classical and instrumental learning by neural networks," *in: Progress in Theoretical Biology, N.Y.: Academic Press*, 1974, 3, pp. 51-141. DOI: 10.1007/978-94-009-7758-7_3

[17] J.C. Baras, and A. La Vigna, "Convergence of the vectors in Kohonen's learning vector quantization," in *Proc. International Neural Network Conference, San Diego, CA,* 1990, pp. 1028-1031. DOI: 10.1007/978-94-009-0643-3_176

[18] Ye. Bodyanskiy, P. Mulesa, O. Slipchenko, and O. Vynokurova, "Self-organizing map and its learning in the fuzzy clustering-classification tasks," *Bulletin of Lviv Polytechnic National University. Computer Science and Information Techology, Lviv: Publishing of Lviv polytechnic,* 2014, 800, pp. 83-92.

[19] T.V. Marushko, "Reactive arthropathy in children," *Health of Ukraine*, 2012, 2, pp. 43-44. (In Ukrainian).

[20] V.M. Savvo, and Yu.V. Sorokolat, "Reactive arthritis in children, which connected with nasopharyngeal infection," *International Medical Journal*, 2003, 9(2), pp. 128-131. (in Russian).

[21] T. Hannu, "Reactive arthritis," *Best Practice & Research Clinical Rheumatology*, 2011, 25, pp.347-357. DOI: 10.1016/j.berh.2011.01.018

[22] C. Selmi, and M.E. Gershwin, "Diagnosis and classification of reactive arthritis," *Autoimmunity Reviews*, 2014, 13, pp.546-549. DOI: 10.1016/j.autrev.2014.01.005

[23] M. Rihl, A. Klos, L.Kohler, and J.G. Kuipers, "Reactive arthritis," *Best Practice & Research Clinical Rheumatology*, 2006, 20(6), pp.1119-1137. DOI: 10.1016/j.berh.2006.08.008

[24] P.S. Kim, T.L. Klausmeier, and D.P. Orr, "Reactive arthritis," *Journal of Adolescent Health*, 2009, 44, pp. 309-315. DOI: 10.1016/j.jadohealth.2008.12.007

Authors' Profiles

Yevgeniy Bodyanskiy. graduated from Kharkiv National University of Radio Electronics in 1971. He got his PhD in 1980. He obtained an academic title of the Senior Researcher in 1984. He got his Dr.habil.sci.ing. in 1990. He obtained an academic title of the Professor in 1994. Prof. Bodyanskiy has been the professor of Artificial Intelligence Department at Kharkiv National University of Radio Electronics, the Head of Control Systems Research Laboratory at Kharkiv National University of Radio Electronics. He has more than 500 scientific publications including 40 inventions and 10 monographs. His research interests are hybrid systems of computational intelligence: adaptive, neuro-, wavelet-, neo-fuzzy-, real-time systems that have to do with control, identification, and forecasting, clustering, diagnostics and fault detection.

Prof. Ye. Bodyanskiy is an IEEE Senior Member and a member of 4 scientific and 7 editorial boards.

Olena Vynokurova graduated from Kharkiv National University of Radio Electronics in 2002. She got her PhD in 2005. She obtained an academic title of the Senior Researcher in 2007. She got his Dr.habil.sci.ing. in 2012. She obtained an academic title of the Professor in 2014. Prof. Vynokurova has been the Principal Research Scientist of Control Systems Research Laboratory at Kharkiv National University of Radio Electronics and the professor of Information Technology Security Department at Kharkiv National University of Radio Electronics. She has more than 150 scientific publications including 4 monographs. Her research interests are hybrid systems of computational intelligence: dynamical data mining, wavelet-neuro-, neo-fuzzy-, on-line systems that have to do with control, identification, and forecasting, clustering, diagnostics and fault detection.

Volodymyr Savvo graduated from Kharkiv National Medical University in 1974. He got his PhD in 1979. He obtained an academic title of the associate professor in 1990. Dr. Savvo has been the associate professor of Pediatrics Department at Kharkiv Medical Academia Post-Graduate Education. He has more than 300 scientific publications.

His research interests are pediatric cardiorheumatology, pediatric gastroenterology, reactive arthritis.

Tetiana Tverdokhlib graduated from Kharkiv National Medical University in 2005. She is a PhD student in Pediatrics at Kharkiv Medical Academia of Post-Graduate Education. Her current interests are reactive arthritis, pediatrics.

Pavlo Mulesa graduated from Kharkiv National University of Radio Electronics in 2008. He got his PhD in Computer Science in 2013. He is currently working as a Senior Researcher at Control Systems Research Laboratory at Kharkiv National University of Radio Electronics. His current research interests are Hybrid Neuro-Fuzzy Systems.