Fuzzy Clustering of Biomedical Datasets Using BSB-Neuro-Fuzzy-Model

Iryna Perova[1[0000-0003-2089-5609]], Yevgeniy Bodyanskiy[1 [0000-0001-5418-2143]], Anatoliy Sachenko[2,3 [0000-0002-0907-3682]], Mikolaj Karpinski[4], Pawel Rudyk[4]

[1] Kharkiv National University of Radio Electronics,
Kharkiv, 61166, Ukraine,
rikywenok@gmail.com, yevgeniy.bodyanskiy@nure.ua
[2] Kazimierz Pulaski University of Technology and Humanities in Radom, Poland
sachenkoa@yahoo.com
[3] Ternopil National Economic University
3 Peremoha Square, Ternopil 46020 Ukraine
as@tneu.edu.ua
[4] University of Bielsko-Biala, 2 Willowa St, 43-309 Bielsko-Biala, Poland,
mkarpinski@ath.bielsko.pl,vagrant326@gmail.com

Abstract: A special neural networks that contain autoassociative memory (AM) - BSB- and GBSB-models are investigated at this paper. These models are implemented on hypercube and solve the task of dataset clusterization due to the fact of point attraction properties of hypercube peaks. A BSB-neuro-fuzzy model can be based on BSB-model as well due to the introduction of the special fuzzy membership function. A training algorithm for the BSB- neuro-fuzzy model is proposed. This algorithm enables to enrich the BSB-neuro-fuzzy model by adaptive properties. An experiment based on medical datasets proved a high quality of the proposed model.

Keywords: fuzzy clustering, hypercube, attractor, adaptive learning algorithm, stable states.

1 Introduction

One of important properties of human brain is a property of information storage and its recovering using association system. Any images ever saw by person can be recovered after long time even in the case of its changing. These brain properties can be simulated by neural networks of associative memory (NN_AM)[1-6].

This artificial memory can be presented by direct-driven neural network (called static associative memory) and by recurrent neural network (called dynamic associative memory), those during its training can stack all patterns (memorizing phase). In recall phase of its functioning dynamic associative memory can make association new proposed pattern with all ever saws. Having said so all patterns ever proposed to the associative memory compose a fundamental memory set.

The basic difference of neural networks of associative memory from approximating neural networks (ANN) consist of the fact that ANN realize nonlinear mapping

$$R^m \ni y(k) = F(x(k) \in R^n),$$

when neural networks of associative memory form mapping of all possible input vectors x in $y(k)$. Input vector x belong to neighbor $x(k)$ such that

$$\|x - x(k)\| < \varepsilon,$$

where $y(k)$ – $(m \times 1)$ fundamental memory vector,

$x(k)$ – $(n \times 1)$ fundamental memory vector,

$k = 1, 2, 3, \ldots, l$ – a total number of fundamental memory pattern,

ε – a special positive parameter.

We are exploring the modified special class of neural network of associative memory is investigated. This memory implements mapping

$$R^n \ni x(k) = F(x(k) \in R^n)$$

for all x belonging to a neighbor area that can be described by ε parameter. The main goal of associative networks is recovering of damaged information or information presented by partial pieces, for example in the area of medical diagnostics when the data fed into processing with gaps and outliers.

2 BSB-neuro model

"Brain-State-In-a-Box Model" was described by D. Anderson with colleagues [7,8]. This model is one of simple and effective architecture amount the structures of associative memory neural network [9-16] and it has the serious theoretical justification.

BSB-model is a neurodynamic nonlinear feedback system with amplitude constraint with the positive feedback. A dynamics of this system can be described in the state space using equation

$$\begin{cases} y(k,\tau) = x(k,\tau) + \alpha W x(k,\tau), \\ x(k,\tau+1) = \psi(y(k,\tau)), \end{cases} \tag{1}$$

where $x(k,0) \equiv x(k)$ – input vector-image; $\tau = 0, 1, 2, \ldots, \mathrm{T}$ – iteration of machine time; $x(k, \mathrm{T})$ – state vector in steady mode; α – small positive parameter of feedback connection; W – $(n \times n)$ – matrix of synaptic weights for correlation AM presented by one-layer neural network formed by adaptive linear associators; $\psi(\bullet)$ – activation piecewise linear function with saturation acting to elements of vector $y(k,\tau)$ component-wise like

$$x_i(k,\tau+1)=\psi(y_i(k,\tau))=\begin{cases}+1, & \text{if} \quad y_i(i,\tau)>+1,\\ y_i(k,\tau), & \text{if} \quad -1\le y_i(k,\tau)\le +1,\\ -1, & \text{if} \quad y_i(k,\tau)<-1,\\ i=1,2,\dots,n. \end{cases} \tag{2}$$

Therefore, the phase space of BSB-model is limited by n-dimensional hypercube whose center is in grid origin and which edge has a length equal two. Whole hypercube has 2^n corners, which should be numbered. For this purpose it is useful to replace negative coordinates by zeros, and after that to change the obtained binary value to the decimal form adding a unit to it. Having said so to corner with all negative coordinates (-1,-1,…,-1) correspond 1-st number and to corner with all positive coordinates (1,1,…,1) – 2^n number.

3 BSB-neuro-fuzzy-model

The BSB-model solves the task of clusterization of input dataset $x(k)$, $k=1,2,\dots,l$ being in same time an AM. All hypercube corners proceed like pointed attractors with a significant domain of attraction to divide all n-dimensional feature space. At this situation a problem connected to capacity W of AM is appearing. The capacity W cannot exceed n (absolute capacity $l/n\le 1$) when the number of hypercube corners equals $2^n >> n$. It is lead to two possible situations: at first, many corners will be «empty» and, at second, data belong to the same cluster can be placed in closely-spaced corners. That's why it is rational to add a special neighborhood function between hypercube corners and consider a pattern from closely-spaced corners belonging to one cluster.

It will be useful to employ ideas of fuzzy clustering [17-21] to apply like neighborhood function the most simple triangle activation function. The membership level of pattern $x(k,\mathrm{T})$ to q-th corner can be defined as

$$\mu_q(x(k,\mathrm{T}))=1-\frac{d(x(k,\mathrm{T}),x_q)}{2n}, \tag{3}$$

where $d(x(k,\mathrm{T}),x_q)=\sum_{i=1}^{n}\left|x_i(k,\mathrm{T})-x_{q,i}\right|$ – Hamming distance between $x(k,\mathrm{T})$ and hypercube corner xq, q=1,2,…,2n. It is easy to see that $\mu_p(x(k,\mathrm{T}))=1$, and membership level for most long-distance corner from x_p^* is equal to zero.

It makes a sense to find the connection between fuzzy clustering based on BSB-model and the most popular fuzzy c-means algorithm (FCM) [17]. In FCM-algorithm the membership level $x(k,\mathrm{T})$ to q-th corner-centroid of cluster can be defined as

$$\mu_q(x(k,\mathrm{T})) = \frac{d^{-1}(x(k,\mathrm{T}),x_q)}{\sum_{l=1}^{2^n} d^{-1}(x(k,\mathrm{T}),x_l)}, \tag{4}$$

where $d(x(k,\mathrm{T}),x_q) = \sum_{i=1}^{n} \left| x_i(k,\mathrm{T}) - x_{q,i} \right|$ – Hamming distance between pattern $x(k,\mathrm{T})$ and hypercube corner xq, q=1,2,…,2n. It interesting to remark that because the pattern $x(k,\mathrm{T})$ belongs to one of hypercube corners, for example, x_p, Hamming distance between x_p and x_q can be defined as double number mismatched coordinate signs, that correspond to these corners.

Equation for $\mu_p(x(k,\mathrm{T}))$ (4) can be transformed to the form

$$\mu_q(x(k,\mathrm{T})) = \frac{d^{-1}(x(k,\mathrm{T}),x_q)}{d^{-1}(x(k,\mathrm{T}),x_q) + \sum_{\substack{l=1 \\ l\neq q}}^{2^n} d^{-1}(x(k,\mathrm{T}),x_l)} =$$

$$= \frac{1}{1 + d(x(k,\mathrm{T}),x_q)\sum_{\substack{l=1 \\ l\neq q}}^{2^n} d^{-1}(x(k,\mathrm{T}),x_l)} =$$

$$= \frac{1}{1 + \dfrac{d(x(k,\mathrm{T}),x_q)}{\sum_{\substack{l=1 \\ l\neq q}}^{2^n} \left(d^{-1}(x(k,\mathrm{T}),x_l)\right)^{-1}}} = \frac{1}{1 + \dfrac{d(x(k,\mathrm{T}),x_q)}{\sigma_q}},$$

where σ_q – a width parameter of bell-shaped membership function of pattern $x(k,\mathrm{T})$ to corner x_q. It is easy to see that if $x(k,\mathrm{T}) \equiv x_q$, membership level is identical equal to one too. The form of membership function for different numbers n of input feature vector $x(k)$ is presenting on Fig.1.

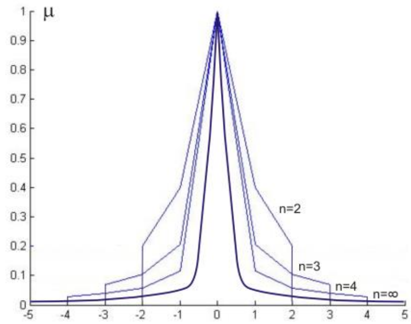

Fig. 1. Membership function of BSB-model

4 Training algorithm of BSB-model

A quality of BSB-model can be defined as a capacity of AM in the feedback circuit. This capacity depends on tuning procedure of n^2 synaptic weights in adaptive linear associators. As such simplest procedure the D. Andersson assessment [7,8] can be used like:

$$W = \sum_{k=1}^{l} x(k)x^{\mathrm{T}}(k) = XX^{\mathrm{T}}, \tag{5}$$

where $X = X(l) = (x(1), x(2), \dots, x(l))^{\mathrm{T}}$ – fundamental memory matrix (matrix size is $(n \times l)$).

In the following we will use also matrixes $X(k) = (x(1), x(2), x(3), \dots, x(k))^{\mathrm{T}}$ $(k < l)$ and $X(1) = x(1)$, excepting $X = X(l)$.

Equation (5) can be rewrited in recurrent form:

$$W(k) = W(k-1) + x(k)x^{\mathrm{T}}(k), \quad W(0) = 0 \cdot \mathrm{I}. \tag{6}$$

It is easy to see that for previously centered and normalized vector $x(k)$ the equation (5) describes autocorrelation matrix on pattern sequence and the expression (6) is a learning Hebb's rule in standard form, widely used in neural networks applications.

For reducing an influence of disturbance component and for improving a quality of recovering we need to minimize errors of recovering $\|v(r)\|$ that means making an orthogonal projection of pattern to fundamental memory vectors. A solving can be obtained after minimizing the criterion

$$E^k = \sum_{k=1}^{l} E(k) = \sum_{k=1}^{l} \|x(k) - Wx(k)\|^2$$

or, it is the same, minimizing the spherical norm

$$E^k = \left\| X - WX \right\|^2 = Tr(X - WX)(X - WX)^{\mathrm{T}} . \tag{7}$$

For these tasks it's expedient to employ the linear projective adaptive algorithms, especially the most widespread autoassociative Widrow-Hoff rule

$$\begin{aligned} W(k) &= W(k-1) + \eta(k)(x(k) - W(k-1)x(k))x^{\mathrm{T}}(k) = \\ &= W(k-1) + \eta(k)(x(k) - \hat{x}(k))x^{\mathrm{T}}(k) \end{aligned} \tag{8}$$

that minimizes criterion

$$E(k) = \left\| x(k) - Wx(k) \right\|^2 ,$$

where $\eta(k)$ is a scalar training rate value which can be selected empirically.

An optimization for a timing this rule leads to a procedure

$$W(k) = W(k-1) + \eta(k) \frac{x(k) - W(k-1)x(k)}{\left\| x(k) \right\|^2} x^{\mathrm{T}}(k) =$$

$$= W(k-1) + \eta(k)\Big(x(k) - W(k-1)x(k)\Big)x^{+}(k) , \; 0 < \eta(k) < 2 ,$$

that can be named is general version of S. Kaczmarz algorithm [22,23] on multidimensional case.

5 GBSB-neuro-model

Nowadays different modifications along with the standard BSB-model (1) are used widely. Among them we may mark out a model [4]

$$x(k, \tau + 1) = \psi(\beta x(k, \tau) + \alpha W x(k, \tau) + \gamma x(k, 0)) ,$$

where $0 < \beta < 1$ – forgetting factor,

γ – a small positive parameter, providing a permanent presence in the model of a pattern $x(k) = x(k, 0)$, that was stored. This modification of BSB-model has high convergence speed and fault tolerance.

At [13,15-16] Generalized Brain-State-in-a-Box Model (GBSB-model) was introduced. The synaptic weights matrix in this model is nonsymmetrical. The dissymmetry can appear after using for Kaczmarz-Widrow-Hoff algorithm and it leads to misconvergence to minimum of adopted energetic function. From other point of view symmetric properties of W accumulate «negatives» of fundamental memory patterns [13] in BSB-model that forms false attractors.

To prevent this disadvantage it is possible to introduce the GBSB-model, using expression:

$$x(k, \tau + 1) = \psi(x(k, \tau) + \alpha W x(k, \tau) + \alpha g) \tag{9}$$

that minimizes the energetic function

$$E = -\left(\frac{\alpha}{2} x^{\mathrm{T}} W x + \alpha x^{\mathrm{T}} g \right),$$

where $g - (n \times 1)$ – a vector, added in (9) for removing false attractors.

6 Experimental results

To investigate BSB-models functioning on medical dataset we have selected a dataset that consists of 182 patterns (patients), each of them is characterizes by 24 features. This dataset describes a psychophysiological human state needed to learn excitative and inhibitory processes on human body. All patients was divided on 2 classes: humans with predominance of excitative processes and ones who are prone to inhibitory processes. All data previously were normalized and centered to be occurred to hypercube $\left[-1;\ 1\right]^n$ and the class feature was eliminated from dataset [24].

All datasets were transmitted to processing on BSB-model and 82% of pattern occur to 2 different corners of hypercube, when other 18% occur to the nearest ones. We have used the membership function (4) to refine the class type of these patterns. Then we have compared the result with known class type and obtained the clusterization accuracy about 92%. The comparison BSB-model with k-means algorithm shows the comparable accuracy of those approaches (about 85% for k-means). The fuzzy c-means algorithm can not be used for a comparison because we need to make the features compression before its clusterization (because of norm effect concentration).

7 Conclusion

The issue of synthesizing adaptive training algorithms for a special type of AM based on BSB- and GBSB-neuro-fuzzy models is considered. The introduced recursive procedures have a high speed, and fuzzy membership functions allows to relate the reconstruction process in the neural network model to the fuzzy-clustering procedures. This approach permits to expands the functionality of the developed method. The practical problem of partitioning into groups (clustering) of factors determining the predominance of excitation/inhibition processes in the body is solved.

References

1. Graupe, D.: Principles of Artificial Neural Networks (Advanced Series in Circuits and Systems). Singapore: World Scient. Publ. Co. Pte. Ltd., (2007).
2. Rojas, R.: Neural Networks. A Systematic Introduction. Berlin: Springer-Verlag (1996).
3. Hagan, M.T., Demuth, H.B., Beale, M.: Neural Network Design. Boston: PWS Publishing Company, (1996).
4. Haykin, S.: Neural Networks and Learning Machine. New York: Prentice Hall, Inc. (2009).

5. K.-L., Du, & M. Swami, Neural Networks and Statistical Learning. London: Springer-Verlag (2014).
6. Bodyanskiy Yevgeniy, Perova Iryna, Fast medical diagnostics using autoassociative neuro-fuzzy memory // Int. J. of Computing, 16 (1), 2017, 34-40.
7. Anderson, J. A.: Cognitive and psychological computation with neural models. IEEE Trans. on Systems, Man, and Cybernetics, (13) 799-815 (1983).
8. Anderson, J.A., Silverstein, J.W., Ritz, S.A., Jones, R.S.: Distinctive features. Categorical perception and probability learning: Some applications of a neural model. In Anderson, J.A., Rosenfeld, E. (eds) "Neurocomputing: Foundations of Research". Cambridge, MA: MIT Press, 413-451 (1988).
9. K. Abe, F. Sevrani, On the Synthesis of Brain-State-in-a-Box Neural Models with Application to Associative Memory. Journal Neural Computation, Vol. 12 (2), 451-472 (2000).
10. Yen, G., Michel, A.N.: A learning and forgetting algorithm in associative memories: Results involving pseudoinverses. IEEE Tr. on Circuits System, (38) 1193-1205 (1991).
11. Zak, S.H., Hui, S., Dynamical analysis of the Brain-State-in-a-Box (BSB) neural models. IEEE Trans. on Neural Networks, #3, 86-94 (1992).
12. Anderson, J.A.: The BSB model: A simple nonlinear autoassociative neural network. In Hassoun M.H. (eds) "Associative neural memories: theory and implementation" N.Y.: Oxford University Press, 77-103 (1993).
13. Lillo, W. E., Miller, D.C., Hui, S., Zak, S.H.: Sythesis of Brain-State-in-a-Box (BSB) based associative memories. IEEE Transactions on Neural Networks, (5) 730-732 (1994).
14. Hui, S., Bohner, M. Brain State in a Convex Body. IEEE Tr. on Neural Networks, (6) 1053-1060 (1995).
15. Park, J., Cho, H., Park, D. Design of GBSB neural associative memories using semidefinite programming IEEE Transactions on Neural Networks, #10 (4), 946-950 (1999).
16. Hui, S., Lillo, W. E., Zak, S.H., , Learning and forgetting in generalized Brain-State-in-a-Box (BSB) neural associative memories. Neural Networks, #9 (5), 845-854 (1996).
17. Bezdek, J.C.: Pattern Recognition with Fuzzy Objective Function Algorithms. N.Y.: Plenum Press, (1981).
18. Kruse, R., Borgelt, C., Klawonn, F., Moewes, C., Steinbrecher, M., Held, P. Computational Intelligence. Methodological Introduction. Berlin: Springer-Verlag, (2013).
19. Kacprzyk, J., Pedrycz, W., Eds., Springer Handbook of Computational Intelligence. Springer-Verlag Berlin Heidelberg, (2015).
20. L.C. Jain, C.L. Mumford Computational Intelligence, Berlin: Springer-Verlag, (2009).
21. Iryna Pliss, Iryna Perova Deep hybrid System of Computational Intelligence with Architecture Adaptation for Medical Fuzzy Diagnostics. I.J.I.S.A., #7, 12-21, (2017).
22. Kaczmarz, S.: Angenaeherte Ausloesung von Systemen linearer Gleichungen. Bull. Int. Acad. Polon. Sci., (Let.A) 355-357 (1937).
23. S. Kaczmarz Approximate solution of linear equations. International Journal Control, #53, 1269-1271, (1993).
24. Yevgeniy Bodyanskiy, Iryna Perova, Adaptive human machine interaction approach for feature selection-extraction task in medical data mining. Int. Journal of Computing, 17 (2), 113-119 (2018).