Neo-Fuzzy System with Special Type Membership Functions Adaptation and Fast Tuning of Synaptic Weights in Emotion **Recognition Task**

Yevgeniy Bodyanskiy, Nonna Kulishova and Olha Chala

Kharkiv National University of Radio Electronics, Nauky Ave. 14, Kharkiv, 61166, Ukraine

Abstract

The neo-fuzzy system for image recognition (by the example of emotion recognition) is proposed. It is designed to solve the task under consideration in conditions of a short dataset and overlapping classes. The distinctive feature of the system is its hybrid learning, which includes controlled learning with a teacher, lazy learning on the principle of "neurons at data points" and self-learning according to T. Kohonen. Also, the ability to tune both synaptic weights and membership functions of special form and provide improved approximation abilities. The system under consideration has a high learning speed and provides good recognition quality that is proved by the results of the computational experiment.

Keywords¹

Neo-fuzzy neuron, nonlinear synapse, controlled learning, lazy learning, self-learning, the adaptive membership function of Epanechnikov kernel type

1. Introduction

Emotions are a powerful tool for interaction between people, which is now increasingly used in various fields of human-computer interaction. The most promising areas of human facial expressions automatic recognition include, for example, education, digital marketing, automated ranking and recommendation systems [1–6]. However, such applications put forward special requirements for recognition approaches: high performance, high accuracy under conditions of significant changes in posture, lighting, and shooting angle. Systems for user's emotional status automatic recognition, as a rule, have a similar architecture, which includes subsystems for pre-processing, feature extraction and facial expression classification. Each of these subsystems can use different approaches to initial data acquisition, machine learning and computational intelligence methods. A wide spectrum of research in this direction has been considered in several reviews [7–9]. To solve the classification problem in emotion detection systems, various machine-learning methods were used [8,10,11].

Since the problem is data-driven, classification accuracy depends on the quality of solutions of previous stages – pre-processing, feature extraction. Deep neural networks permit to improve the recognition accuracy significantly [9]. Among similar architectures Convolutional Neural Networks (CNN), attentional CNN [12,13], graph CNN [14–16], component-wise LSTM (cLSTM) [17] and many others were proposed. Despite the variety of deep networks applied to facial expression recognition, they all have serious common disadvantages. First, the recognition accuracy of trained networks strongly depends on how diverse and large the dataset to network learning was used, whether it contained information about representatives of different races, ages, and cultures. When preparing such datasets, specialists conduct video or photography in studio conditions, when a person's posture, movements, nature, lighting almost do not change, and emotions are manifested as much as possible for an unambiguous interpretation. These factors reduce adaptive recognition

EMAIL: yevgeniy.bodyanskiy@nure.ua (A. 1); nonna.kulishova@nure.ua (A. 2); olha.chala@nure.ua (A. 3) ORCID: 0000-0001-5418-2143 (A. 1); 0000-0001-7921-3110 (A. 2); 0000-0002-7603-1247 (A. 3)



^{© 2021} Copyright for this paper by its authors. Use permitted under Creative Commons License Attribution 4.0 International (CC BY 4.0).

CEUR Workshop Proceedings (CEUR-WS.org)

II International Scientific Symposium «Intelligent Solutions» IntSol-2021, September 28–30, 2021, Kyiv-Uzhhorod, Ukraine

capabilities of trained deep network. In addition, dataset formation and labeling required for deep network training is time and labor costly, and dataset size can reach millions of samples.

These aspects significantly limit possibilities of deep networks used for real-time recognition, when emotions are manifested to a small extent, mixed, while training data volumes are small, and samples can be unlabelled. Therefore, the problem of developing a system for automatic emotions recognition in real time remains actual. One of the helpful approaches is the use of neo-fuzzy neurons based systems [18], where the solution of the fuzzy classification problem [19,20] is implemented.

In this work, it is also proposed to consider a person's facial expression recognition as a fuzzy classification problem, and to use it for solving a modification of a neo-fuzzy neuron with membership functions such as Epanechnikov kernel with tuning centers.

2. Architecture of neo-fuzzy system for emotion recognition

Among many approaches, deep neural networks are adjusted in the best way to solve pattern recognition task. Such systems proved their effectiveness in solving many problems, which are related to the procession of big amount of information. At the same time, these neural networks are quite slow, contain huge among of tuning synaptic weights (sometimes billions, some – more than trillion). Hence, they require huge amount of training data for their learning. In case the size of training dataset is limited that appends often in real tasks, deep neural networks become noneffective and the transfer learning usage is not always allow solve arising problems.

A neuro-fuzzy neuron [21–25] is the simplest system that allows restoring separating hypersurfaces proved its effectiveness for solving numerous real tasks. In general case, this construction is Takagi-Sugeno-Kang system of a zeroth-order differential equation, viz, it is a universal approximator. The architecture of standard neo-fuzzy neuron is represented in Fig.1.

Input vector-image $x(k) = (x_1(k), ..., x_i(k), ..., x_n(k))^T \in \mathbb{R}^n$ (here k = 1, 2, ..., N – or a number of an observations from a training dataset, or current discrete time) is fed to the inputs of nonlinear synapse NS_i each of which contains h_i membership functions $\mu_{li}(x_i)$ and same number of tunable synaptic weights w_{li} that are specified or in batch mode or in online mode through optimisation of the adopted learning criteria – goal function.

In the general case, neo-fuzzy neuron implements nonlinear transformation:

$$\hat{y}(k) = \sum_{i=1}^{n} f_i(x_i(k)) = \sum_{i=1}^{n} \sum_{l=1}^{n_i} \mu_{li}(x_i(k)) w_{li}(k-1),$$
(1)

thereat due to the linear dependence between output signal $\hat{y}(k)$ and synaptic weights, the algorithms of linear adaptive identification, including optimal by speed, robust, with smoothing [25–27] can be used for neo-fuzzy neuron's learning.

Usually in neo-fuzzy neurons B-splines are used as an activation functions, because they satisfy Ruspini unity partition conditions. Thereby, neo-fuzzy neuron does not contain defuzzification layer, that takes a lot of hard work out of its implementation. Typically there are the first-order B-splines in other words they are traditional triangular membership functions. Their main advantage is to fire only two neighboring functions in discrete moment k, and in each nonlinear synaps only two neighboring synaptic weights are tuned (total x in neo-fuzzy neuron 2n). Such approach essentially improves learning speed, especially when data are fed to the processing sequentially in online mode.

Meanwhile, such neo-fuzzy neuron can implement only piecewise approximation of separating hypersufraces that cannot always provide necessary quality of the classification. In this context in [28] the extended neo-fuzzy neuron was introduced, where each non-linear signal of each neuron is Takagi-Sugeno-Kang fuzzy system of an arbitrary orderfuzzy. The image recognition system is based on the extended neo-fuzzy neurons, provides high quality of the solution of the classification task, however, it still require increased in size training datasets, because it has considerably larger number of tunable weights. Hence it is reasonable to use other kernel functions than B-splines in standart neo-fuzzy neuron, and the simplest examples is Epanechnikov kernels [29] that are represented in Fig.2.





Here $x_{i\min}$, $x_{i\max}$ interval of input signal adjustment on *i*-th input. If on this input h_i such functions are evenly allocated, then interval between neighbouring centers is set with the formula

$$r_i = \frac{x_{i\min} - x_{i\max}}{h_i - 1}.$$
 (2)

These functions can be written in analytical form:

$$\mu_{li}(x_i) = (1 - (x_i - c_{li})^2 r_i^{-2}) \delta_{li}$$
(3)

where c_{li} - centers of corresponding functions,

$$\delta_{li} = \begin{cases} 1 & if |x_i - c_{li}| < r_i, \\ 0 & otherwise. \end{cases}$$
(4)

In more general case membership functions can be distributed nonuniformly as it is shown in Fig.3.



Figure 3: Unsymmetrical membership functions of Epanechnikov kernel type.

It is readily seen that in this situation membership functions are unsymmetrical and can be described with the following equitation:

$$\begin{cases} \mu_{liL}(x_i) = \left[1 - \frac{(x_i - c_{li})^2}{(c_{l-1,i} - c_{li})^2}\right]_+, \\ \mu_{liR}(x_i) = \left[1 - \frac{(x_i - c_{li})^2}{(c_{l+1,i} - c_{li})^2}\right]_+ \end{cases}$$
(5)

where $[\Box]_+$ - a projection on positive ortant.

It is easy to see that i each moment k, only two neighbouring functions can be fired as well, their derivatives are equal to zero in centers. However, these functions do not satisfy Ruspini unity partition conditions, in other words the system is built on the basis of such neo-fuzzy neurons requires additional output defuzzification layer. Having used neo-fuzzy neuron, the binary classification task can be solved, however, inasmuch as this case it is nessesary to split initial dataset to m possibly overlapping classes. Thus it is reasonable to introduce neo-fuzzy system, that is designed to solve pattern recognition task, and architecture of which is represented in Fig. 4. The system contains m connected in parallel neo-fuzzy neurons, where their outputs $y_j(k)$, j = 1, 2, ..., m are formed with softmax activation functions, that are usually form output signals of deep convolutional neural network, that solve classification task. Therefore on the outputs of the neo-fuzzy system signals are formed:

$$y_{j}(k) = \operatorname{softmax} \hat{y}_{j}(k) = \frac{e^{\hat{y}_{j}(k)}}{\sum_{i=1}^{m} e^{\hat{y}_{j}(k)}}$$
 (6)



Figure 4: Neo-fuzzy system for image recognition

whereas $\sum_{j=1}^{m} e^{\hat{y}_j(k)} = 1$. These signals set membership level of the observation x(k) to the *j*-th class.

It is interesting to notice that output softmax layer play the role defuzzification layer in neurofuzzy systems that is to say in this system it is not necessary for membership functions to meet requirements of the unity partitioning.

3. Learning of neo-fuzzy system

The cross-entropy, which is usually used for deep convolutional neural networks tuning, underlies the learning criteria that is used for the learning of the proposed neo-fuzzy system:

$$E(k) = \sum_{j=1}^{m} E_j(k) = -\sum_{j=1}^{m} y_j^*(k) \ln y_j(k)$$
(7)

where $y_j^*(k)$ – external reference signal, which takes only two values: 1 if the vector-image x(k) belongs to the *j*-th class and 0 otherwise.

Let us introduce into consideration vectors of synaptic weights and membership functions of *j*-th neo-fuzzy neuron, that have following dimension: $\left(\sum_{i=1}^{N} h_i \times 1\right)$: $\mu_j(x) = (\mu_{j11}(x_1), ..., \mu_{jh1}(x_1), \mu_{j12}(x_2), ..., \mu_{jh2}(x_2), ..., \mu_{jli}(x_i), ..., \mu_{jhn}(x_n))^T$, $w_j = (w_{j11}, ..., w_{jh1}, w_{j12}, ..., w_{jh2}, ..., w_{jli}, ..., w_{jlin})^T$ thus its output signal can be written in the following form:

$$\hat{y}_{j}(k) = w_{j}^{T}(k-1)\mu_{j}(x(k)).$$
 (8)

Introducing vector reference signal $y^*(k) = (y_1^*(k), ..., y_j^*(k), ..., y_m^*(k))^T$ which is formed with zeroes and ones (so-called "one-hot coding"), vector output signals of the system in general $y(k) = (y_1(k), ..., y_j(k), ..., y_m(k))^T$ and $\hat{y}(k) = (\hat{y}_1(k), ..., \hat{y}_j(k), ..., \hat{y}_m(k))^T$, $(m\sum_{i=1}^n h_i \times 1)$ vector of membership functions $\mu(x(k)) = (\mu_1^T(x(k)), ..., \mu_j^T(x(k)), ..., \mu_m^T(x(k)))^T$ and $(m \times m\sum_{i=1}^n h_i)$

matrix of synaptic weights

$$w(k-1) = \begin{pmatrix} w_1^T (k-1) \\ \vdots \\ w_j^T (k-1) \\ \vdots \\ w_m^T (k-1) \end{pmatrix}$$
(9)

the output signals of the system can be written in vector-matrix form:

$$\hat{y}(k) = w(k-1)\mu(x(k)),$$
(10)

$$y(k) = \frac{e^{\hat{y}(k)}}{I^T e^{\hat{y}(k)}} = \frac{e^{w(k-1)\mu(x(k))}}{I^T e^{w(k-1)\mu(x(k))}}$$
(11)

where $I - (m \times 1)$ vector that is formed with unities.

Matrix version of optimal by speed learning algorithm of Kaczmarz-Widrow-Hoff can be used for the tuning of the matrix of synaptic weights w(k) and written in the form:

$$w(k) = w(k-1) + \frac{y^*(k) - w(k-1)\mu(x(k))}{\|\mu(x(k))\|^2} \mu^T(x(k))$$
(12)

or its adaptive regularized modification:

$$w(k) = w(k-1) + \frac{y^{*}(k) - w(k-1)\mu(x(k))}{\gamma + \|\mu(k)\|^{2}} \mu^{T}(x(k))$$
(13)

(here $\gamma > 0$ – momentum term), protected from the "exploding gradient".

The quality of learning can be improved through turning not only synaptic weights, but also centers of membership functions of nonlinear synapse. In order to prevent training dataset from growing it is better to use ideas of selflearning, that are developed by T. Kohonen [30] as well as lazy learning[31]. Let's introduce in consideration certain threshold of indecomposability that set minimal possible distance between neighbouring centers $\mathbf{r}_{i\min} = |c_{li} - c_{l-1,i}|_{\min}$. Then the initial stages such process take place in the following way:

- Feeding nonlinear synapse NS_i on the input of the signal $x_i(1)$, first center c_{1i} is formed;
- Feeding nonlinear synapse signal $x_i(2)$ on the input, the condition is checked:

$$|x_i(2) - c_{1i}| \le r_{i\min};$$
 14)

- In case this condition are met, nothing will happen new centers will not be formed;
- If the following condition is satisfied:

$$r_{i\min} < |x_i(2) - c_{1i}| \le 2r_{i\min}$$
 (15)

centers are corrected according to Kohonen's rule "winner takes all"[30]:

$$c_{1i}(2) = c_{1i}(1) + \eta(2)(x_i(2) - c_{1i}(1))$$
(16)

(here $\eta(2)$ - selflearning rate parameter);

• If the following condition is satisfied:

$$2r_{\min} < |x_i(2) - c_{1i}| \tag{17}$$

according to the lazy learning rule "Neurons at data points", the second center is formed $c_{2i} = x_i(2)$ and formed early centers c_{1i} stay where they are. The process of centers tuning take place until h_i centers will be formed, where this value is defined according to the following formula:

$$h_i = \frac{x_{i\max} - x_{i\min}}{r_{i\min}} + 1.$$
 (18)

163

1

Further, only their coordinates are corrected according to selflearning algorithm. Such tuning of kernel membership functions' location allows to improve the approximation abilities of the system.

4. Results

The accuracy and speed of proposed system were investigated on a dataset that was formed from known Psychological Image Collection at Stirling (PICS) [32], Extended Cohn-Kanade (CK +) databases [33]. The set contains 821 images that convey emotion development in dynamics and also contain micro-facial expressions (Fig. 5). An array of 35 feature points was selected as a face model (Fig. 6).



Figure 5: Photos from dataset showing emotions development over time, including micro-expressions



Figure 6: Location of 35 feature points

All images represent basic emotions: surprise, joy, disgust, grief, anger, fear, and neutral expression. Thus, there are 7 classes in the classification problem.

5. Conclusion

The neo-fuzzy system and its combined learning (controlled learning, self-learning, leaning leaning) were proposed and deigned to solve image-emotion recognition task under the conditions of limited by volume dataset. The main characteristic of the proposed system is the usage a special kernel constructions as a activation functions that allows improving approximation properties of the system. On behalf of the controlled learning, the optimal by speed algorithm adjusted for the conditions of short dataset is used. Additionally, lazy learning and self-leaning allow disposing placing membership functions belongings in nonlinear synapse in the optimal way. The proposed system is quite simple in calculation implementation and provides high quality of recognition that is proved by the computational experiment.

6. References

- [1] F. Alqahtani, N. Ramzan, Comparison and Efficacy of Synergistic Intelligent Tutoring Systems with Human Physiological Response, Sensors, 2019. https://doi.org/10.3390/s19030460.
- [2] Z. Hussain, M. Zhang, X. Zhang, K. Ye, C. Thomas, Z. Agha, N. Ong, A. Kovashka, Automatic Understanding of Image and Video Advertisements, in: 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), IEEE, Honolulu, HI, 2017, pp. 1100–1110. https://doi.org/10.1109/CVPR.2017.123.
- [3] J.J. Sun, T. Liu, G. Prasad, GLA in MediaEval 2018 Emotional Impact of Movies Task, CoRR. abs/1911.12361 (2019). http://arxiv.org/abs/1911.12361.
- [4] W. Hua, F. Dai, L. Huang, J. Xiong, G. Gui, HERO: Human Emotions Recognition for Realizing Intelligent Internet of Things, IEEE Access, 2019, pp. 24321–24332. https://doi.org/10.1109/ACCESS.2019.2900231.
- [5] Z. Wei, J. Zhang, Z. Lin, J.-Y. Lee, N. Balasubramanian, M. Hoai, D. Samaras, Learning Visual Emotion Representations From Web Data, in: 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 2020, pp. 13103–13112. https://doi.org/10.1109/CVPR42600.2020.01312.
- [6] M.S. Hossain, G. Muhammad, An Emotion Recognition System for Mobile Applications, IEEE Access, 2017, pp. 2281–2287. https://doi.org/10.1109/ACCESS.2017.2672829.
- [7] C.A. Corneanu, M.O. Simon, J.F. Cohn, S.E. Guerrero, Survey on RGB, 3D, Thermal, and Multimodal Approaches for Facial Expression Recognition: History, Trends, and Affect-Related Applications 38 volume of IEEE Trans. Pattern Anal. Mach. Intell., 2016 1548–1568. https://doi.org/10.1109/TPAMI.2016.2515606.
- [8] B. Martinez, M.F. Valstar, B. Jiang, M. Pantic, Automatic Analysis of Facial Actions: A Survey, 10 volume of IEEE Trans. Affective Comput., 2019, 325–347. https://doi.org/10.1109/TAFFC.2017.2731763.
- [9] S. Li, W. Deng, Deep Facial Expression Recognition: A Survey, IEEE Trans. Affective Comput., 2020. https://doi.org/10.1109/TAFFC.2020.2981446.
- [10] Z. Zhang, P. Luo, C.C. Loy, X. Tang, From Facial Expression Recognition to Interpersonal Relation Prediction, CoRR., 2016.
- [11] A. Aslam, B. Hussian, Emotion recognition techniques with rule based and machine learning approaches, CoRR., 2021.
- [12] K.-C. Liu, C.-C. Hsu, W.-Y. Wang, H.-H. Chiang, Real-Time Facial Expression Recognition Based on CNN, in: 2019 International Conference on System Science and Engineering (ICSSE), IEEE, Dong Hoi, Vietnam, 2019, pp. 120–123. https://doi.org/10.1109/ICSSE.2019.8823409.
- [13] S. Minaee, A. Abdolrashidi, Deep-Emotion: Facial Expression Recognition Using Attentional Convolutional Network, CoRR., 2019.

- [14] W.-S. Chien, H.-C. Yang, C.-C. Lee, Cross Corpus Physiological-based Emotion Recognition Using a Learnable Visual Semantic Graph Convolutional Network, in: Proceedings of the 28th ACM International Conference on Multimedia, ACM, Seattle WA USA, 2020, pp. 2999–3006. https://doi.org/10.1145/3394171.3413552.
- [15] Y. Fan, J.C.K. Lam, V.O.K. Li, Facial Action Unit Intensity Estimation via Semantic Correspondence Learning with Dynamic Graph Convolution, CoRR., 2020.
- [16] L. Lo, H.-X. Xie, H.-H. Shuai, W.-H. Cheng, MER-GCN: Micro Expression Recognition Based on Relation Modeling with Graph Convolutional Network, CoRR., 2020.
- [17] T. Mittal, P. Mathur, A. Bera, D. Manocha, Affect2MM: Affective Analysis of Multimedia Content Using Emotion Causality, CoRR., 2021.
- [18] Y. Bodyanskiy, N. Kulishova, O. Chala, The Extended Multidimensional Neo-Fuzzy System and Its Fast Learning in Pattern Recognition Tasks, 3 volume of Data, 2018. https://doi.org/10.3390/data3040063.
- [19] Y. Bodyanskiy, O. Chala, I. Pliss, A. Deineko, Adaptive Probabilistic Neural Network With Fuzzy Inference And Its Online Learning, in: 2020 IEEE 15th International Conference on Computer Sciences and Information Technologies (CSIT), IEEE, Zbarazh, Ukraine, 2020, pp. 96–99. https://doi.org/10.1109/CSIT49958.2020.9322052.
- [20] Y. Bodyanskiy, I. Pliss, O. Chala, A. Deineko, Evolving fuzzy-probabilistic neural network and its online learning, in: 10th International Conference on Advanced Computer Information Technologies, Deggendorf, Germany, 2020.
- [21] E. Uchino, T. Yamakawa, Soft Computing Based Signal Prediction, Restoration, and Filtering, in: D. Ruan (Ed.), Intelligent Hybrid Systems, Springer US, Boston, MA, 1997, pp. 331–351. https://doi.org/10.1007/978-1-4615-6191-0_14.
- [22] T. Miki, Analog Implementation of Neo-Fuzzy Neuron and Its On-board Learning, 1999.
- [23] D. Zurita, M. Delgado, J.A. Carino, J.A. Ortega, G. Clerc, Industrial Time Series Modelling by Means of the Neo-Fuzzy Neuron, volume 4 of IEEE Access, 2016 6151–6160. https://doi.org/10.1109/ACCESS.2016.2611649.
- [24] T. Yamakawa, E. Uchino, J. Miki, H. Kusanagi, A neo-fuzzy neuron and its application to system identification and prediction of the system behavior, in: Proceedings of the 2nd International Conference on Fuzzy Logic & Neural Networks, Iizuka, Japan, 1992.
- [25] Y. Bodyanskiy, I. Kokshenev, V. Kolodyazhniy, An adaptive learning algorithm for a neo-fuzzy neuron, in: Proceedings of the 3rd Conference of the European Society for Fuzzy Logic and Technology, Zittau, Germany, 2003.
- [26] Y. Bodyanskiy, S. Popov, M. Titov, Robust Learning Algorithm for Networks of Neuro-Fuzzy Units, in: T. Sobh (Ed.), Innovations and Advances in Computer Sciences and Engineering, Springer Netherlands, Dordrecht, 2010, pp. 343–346. https://doi.org/10.1007/978-90-481-3658-2_59.
- [27] G.C. Goodwin, P.J. Ramadge, P.E. Caines, Discrete Time Stochastic Adaptive Control, 19 volume of SIAM J. Control Optim., 1981, 829–853. https://doi.org/10.1137/0319052.
- [28] Ye.V. Bodyanskiy, N.E. Kulishova, Extended neo-fuzzy neuron in the task of images filtering, Radio Electronics, Computer Science, Control., 2014. https://doi.org/10.15588/1607-3274-2014-1-16.
- [29] V.A. Epanechnikov, Non-Parametric Estimation of a Multivariate Probability Density, volume14 of Theory Probab. Appl., 1969, 153–158. https://doi.org/10.1137/1114019.
- [30] T. Kohonen, Self-Organizing Maps, Springer Berlin Heidelberg, Berlin, Heidelberg, 2001. https://doi.org/10.1007/978-3-642-56927-2.
- [31] D.R. Zahirniak, R. Chapman, Rogers, Pattern recognition using radial basis function networks, in: Sixth Annual Aerospace Applications of AI Conf, Dayton, 1990, pp. 249–260.
- [32] 2D face sets, (n.d.). http://pics.psych.stir.ac.uk/2D_face_sets.htm (accessed July 13, 2021).
- [33] P. Lucey, J.F. Cohn, T. Kanade, J. Saragih, Z. Ambadar, I. Matthews, The Extended Cohn-Kanade Dataset (CK+): A complete dataset for action unit and emotion-specified expression, in: 2010 IEEE Computer Society Conference on Computer Vision and Pattern Recognition -Workshops, 2010, pp. 94–101. https://doi.org/10.1109/CVPRW.2010.5543262.