Artificial Complex Neurons with Half-plane-like and Angle-like Activation Function

Vladyslav Kotsovsky, Fedir Geche, Anatoliy Batyuk

Abstract – **The paper deals with the problems of Boolean functions realization on neural-like units with complex weight coefficients. The relation between classes of realizable function is considered for half-plane-like activation function. We also introduce the concept of sets separability, corresponding to our notion of neuron. The iterative online learning algorithm is proposed and sufficient conditions of its convergence are given. We also consider complex neurons with angle-type activation functions.**

Key words – **complex neuron, neural network, threshold function, activation function, learning.**

I. INTRODUCTION

Artificial neural networks based on neural-like units have numerous applications in different areas, such as artificial intelligence, objects classification, pattern recognition, data compression, forecasting, approximation or extrapolation of functions of many variables and many others [1]. Different networks architectures and neuron kinds are described in [1, 2]. One of most important task in the theory of feed-forward neural networks with discrete activation functions is the one concerning the realization of a Boolean function on a single neuron. Its importance follows from the fact that for networks on the base of neurons with threshold-like activation function outputs of each network levels have two possible values (binary, bipolar, etc.). Minsky and Papert [3] proved that classical threshold units have enough weak capacity for recognition.

Numerous improved models of neuron are proposed for overcome the mentioned limitations (see [1] for details). In paper we deal with the one type of such extensions, namely discrete-valued complex neurons, which were introduced in [4]. There exists many way of complexification, e.g. [4, 5]. Since nineties artificial neural networks based on complex neurons with different kind of continuous activation functions are widely used in complex signal processing [6]. But in many task arising in digital signal processing discrete valued neurons are preferable [5] to the continuous ones.

Vladyslav Kotsovsky — Information managing systems and technologies department, Uzhgorod National University, Pidgirna Str., 46, Uzhgorod, 88000, UKRAINE,
E-mail: Kotsavlad@gmail.com

Fedir Geche — Cybernetics and Applied Mathematics Department, Uzhgorod National University, Pidgirna Str., 46, Uzhgorod, UKRAINE
E-mail: fgeche@hotmail.com

Anatoliy Batyuk — ACS Department, Lviv Polytechnic National University, S. Bandery Str., 12, Lviv, 79013, UKRAINE,
E-mail: abatyuk@gmail.com

II. COMPLEX NEURONS

We consider the following way to extend the notions of neural unit and threshold function [1] to the complex domain. Let us consider Boolean function over alphabet $\{\alpha,\beta\}$ where α and β are complex number. Let l be an arbitrary line dividing the complex plane C on two half-plane C^+ and C^-. We may regard following sign function generated by above mentioned division:

$$\operatorname{sgn}_l z = \begin{cases} \alpha & \text{if } z \in C^-, \\ \beta & \text{if } z \in C^+ \cup l. \end{cases}$$

The graphical illustration of the division of the complex plane is given on Fig. 1.

A Boolean function $f:\{\alpha,\beta\}^n \to \{\alpha,\beta\}$ is a complex Boolean threshold function (CBTF) in the alphabet $\{\alpha,\beta\}$ *if there exists a complex weight vector* $\mathbf{w} \in C^{n+1}$ and line l such that

$$f(\mathbf{z}) = \operatorname{sgn}_l(\mathbf{w}, \overline{\mathbf{z}}'),$$

where $\mathbf{z} = (z_1,\dots,z_n) \in \{\alpha,\beta\}^n$, $\mathbf{z}' = (z_1,\dots,z_n,1)$ $\overline{\mathbf{z}}'$ is a complex conjugate vector for $\mathbf{z}'$ (here we used the definition of inner product in complex vector spaces). Similarly to the real case it is natural to call a vector $\mathbf{w}$ a weight vector of complex neuron.

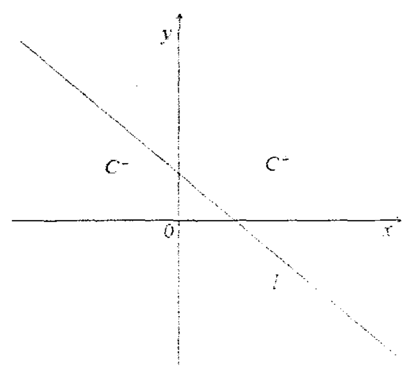

Fig. 1. Surface of sgn_l sign function

Note that we do not use the notion of the threshold in our definition, because it is convenient to include the threshold in the weight vector.

It is easy to see that using rotation and fitting of the free term w_{n+1} we can restrict the class of possible sign function to the following function

$$\operatorname{Resgn} z = \begin{cases} \alpha & \text{if } \operatorname{Re} z < 0, \\ \beta & \text{if } \operatorname{Re} z > 0. \end{cases}$$

Note that "small" change of term w_{n+1} allows avoiding the possibility that the value of weighted sum $(\mathbf{w}, \bar{\mathbf{z}}')$ lies on the division line.

Let $T_{\mathrm{C}}(\alpha,\beta)$ be a class of all CBTF in alphabet $\{\alpha,\beta\}$. The question arises about relations existing among the classes of CBTF in different alphabets. The answer is given by the following proposition.

Proposition 1. *There exists an bijective correspondence between the classes* $T_{\mathrm{C}}(\alpha,\beta)$ *and* $T_{\mathrm{C}}(\gamma,\delta)$ *for arbitrary alphabets* $\{\alpha,\beta\}$, $\{\gamma,\delta\}$.

Note, in particular, that one cannot obtain the class of CBTF more powerful that $T_{\mathrm{C}}(-1,1)$ by altering the alphabet.

The next question is how the cardinality of the class of CBTF changes if we restrict the set of possible value for weight vector coefficients. Let $T_D^n(\alpha,\beta)$ be the class of all CBTF of n variables realizable on neurons with weight vectors from the set D^{n+1},

$$T_D(\alpha,\beta) = \bigcup_{n=0}^{\infty} T_D^n(\alpha,\beta),$$

where $D \subseteq \mathrm{C}$.

Proposition 2. *If* $\operatorname{Re}\alpha \neq \operatorname{Re}\beta$, *then*

$$T_{\mathrm{C}}(\alpha,\beta) = T_D(\alpha,\beta).$$

Note that in the case of bipolar alphabet $E_2 = \{-1,1\}$ the proposition was proved in [4] similar to the last proposition.

From the previous proposition also follows that usage of neurons with weights belonging to the real line enable us to generate all CBTF. We will prove that similar fact is true for neurons with weights lying on any line in complex space.

Proposition 3. *If* $\gamma \in \mathrm{C}$, $\gamma\mathrm{R} = \{\gamma\mathbf{x} \mid \mathbf{x} \in \mathbf{R}\}$ *and complex numbers* α,β,γ *satisfy conditions*

$$|\arg\gamma| < \frac{\pi}{2}, \ \operatorname{Re}(\alpha-\beta)\gamma \neq 0,$$

then classes $T_{\mathrm{C}}(\alpha,\beta)$ and $T_{\gamma\mathrm{R}}(\alpha,\beta)$ *coincide.*

III. LEARNING ALGORITHM

We have seen that $T_{\mathrm{C}}(\alpha,\beta) = T_{\gamma\mathrm{R}}(\alpha,\beta)$, and question how find some weight vector $\mathbf{w} \in T_{\gamma\mathrm{R}}(\alpha,\beta)$, corresponding to given CBTF f naturally arises. That is, we need a learning algorithm for the class of CBTF.

Let A^+, A^- be two finite disjunctive subsets of vectors from the set $\mathrm{C}^n \times \{\gamma\}$. ($\gamma \neq 0$) (i.e. $A^+ \cap A^- = \varnothing$) and $A = A^+ \cup A^-$. We call sets A^+ and A^- γ-separable, if there exists vector $\mathbf{w} \in \gamma R^{n+1}$ such that for all $\mathbf{z} \in A$ following conditions hold

$$(\mathbf{w}, \bar{\mathbf{z}}) > 0 \ \text{ if } z \in A^+,$$
$$(\mathbf{w}, \bar{\mathbf{z}}) < 0 \ \text{ if } z \in A^-.$$

Next, we will suppose that there exists an angle ϕ and real number c such that

$$\forall \mathbf{z} \in A \ \left|\operatorname{Re}\left(e^{i\phi} z_j\right)\right| \geq c > 0 \ \ (j = 1,\ldots,n). \qquad (1)$$

We will assume Eq. (1), without any loss of generality, because A is a finite set.

Let the training sample of vectors $\{\mathbf{z}^k\}$ satisfies following two conditions:

1) $\mathbf{z}^k \in A$, $k \in \mathrm{N}$;

2) each element of the set A repeats in learning sample infinitely many times.

Without any loss of generality we will assume that $\gamma = e^{i\phi}$, where $-\frac{\pi}{2} < \phi < \frac{\pi}{2}$. Let the initial weight vector be $\mathbf{w}^0 = (0,\ldots,0)$. Let us build the sequence of vectors $\{\mathbf{w}^k\}$ as follow:

$$\mathbf{w}^k = \mathbf{w}^{k-1} + t_k h_\phi\left(\mathbf{z}^k\right) e^{i\phi}, \qquad (2)$$

where $h_\phi(\mathbf{z}) = \left(\operatorname{Re}\left(\bar{\mathbf{z}}_1 e^{-i\phi}\right),\ldots,\operatorname{Re}\left(\bar{\mathbf{z}}_n e^{-i\phi}\right),1\right)$, and a coefficient t_k in defined by

$$t_k = \begin{cases} 1 & \text{if } \operatorname{Re}\left(\mathbf{w}^{k-1}, \overline{z^k}\right) \leq 0 \text{ and } z \in A_+, \\ -1 & \text{if } \operatorname{Re}\left(\mathbf{w}^{k-1}, \overline{z^k}\right) \geq 0 \text{ and } z \in A_-, \\ 0 & \text{otherwise.} \end{cases} \qquad (3)$$

The algorithm of weights updating according to the rule (2)-(3) we call "the online learning algorithm" for the complex neural unit. The next proposition gives the sufficient condition for our learning algorithm to be convergent.

Proposition 4. *If finite sets* A^+ *and* A^- *are* γ-*separable, then there exists finite natural m such that the sequence of weight vectors, obtaining according to the* *rules (2)-(3) of online learning algorithm yield after m updates the weight vector* $\mathbf{w}^m$, *which separates sets* A^+ *and* A^-.

IV. Neurons with Angle-Like Activation Function

It follows from proposition 2 that in case of half-plane like activation function complex weights do not provide any advantage in comparison with real weights. Therefore, we need more powerful discrete activation functions.

Let us consider one of many possible generalization of half-plane like activation functions. Let us take some angle on complex plane. Then we can define the sign function of the neuron with bipolar inputs in the following way:

$$\text{Asign } z = \begin{cases} 1, & \text{if } z \text{ falls inside the angle,} \\ -1, & \text{if } z \text{ falls outside the angle.} \end{cases}$$

Note that we can classify points on the sides of angle in any convenient way.

Without loss of generality, we can limit ourselves by following sign functions

$$\text{sign}_{\varphi} z = \begin{cases} 1, & 0 \le \text{Arg } z \le \varphi, \\ -1, & \text{otherwise.} \end{cases}$$

The correspondent decision regions are shown in Fig. 2.

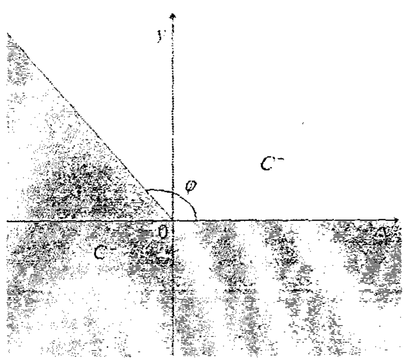

Fig. 2. Angle-like activation function

Let us consider complex neuron with following weights: $w_1 = 1 - i$, $w_2 = -1 + i$, $w_3 = 1 + i$ and let φ be $\frac{\pi}{2}$. Then $z_1 z_2 = \text{sign}_{\frac{\pi}{2}}\left(\mathbf{w}, \overline{\mathbf{z}}'\right)$. Thus, we obtained the function corresponding to the solution of the XOR problem in threshold logic. It follows that neurons with angle-like activation functions are more powerful than classical threshold units.

We propose the following correction rule yielding the perceptron-like learning algorithm:

$$\mathbf{w}^k = \mathbf{w}^{k-1} + t_k e^{i\varphi/2} \mathbf{z}'^k , \qquad (4)$$

where $t_k = 1$ if $\mathbf{z}^k \in C_+$ or $t_k = -1$ if $\mathbf{z}^k \in C_-$, and vector $\mathbf{z}^k$ is misclassified by neuron with weight vector $\mathbf{w}^{k-1}$. In both cases the weighted sum $\left(\mathbf{w}^k, \mathbf{z}'^k\right)$ is "closer" to the necessary part of the complex plane than precedent weighted sum $\left(\mathbf{w}^{k-1}, \mathbf{z}'^k\right)$.

Note that correction rule given by Eq. (4) is very similar to the same in Eq. (2). But rule (4) is quite heuristic in the case of complex neuron with angle-like activation function. Unlike the proposition 4 we have not exact proof of its convergence.

V. Conclusion

Artificial complex neurons with the half-plane surface of activation function are enough simple and powerful computational units. Our main results concerning complex neurons with discrete activation function are following:

1. The choice of the alphabet of Boolean functions representation has no importance for representative power of class of respective realizable Boolean functions.
2. The restriction of possible weights to ones on an almost every line in complex plane does not shrink the class of respective complex Boolean threshold functions.
3. Neurons with restricted weights can be learned by using convergent perceptron-like learning technique.
4. Complex neuron with angle-like activation function can be learned by perceptron-like correction learning rule.

References

[1] S. Haykin, "Neural networks: A comprehensive foundation", Upper Saddle River, NJ: Prentice Hall, Inc., 1999.

[2] C. M. Bishop, "Neural networks for pattern recognition", Oxford University Press, Oxford, UK, 1995.

[3] M. Minsky, S. Papert, "Perceptrons: An introduction to computational geometry". MIT Press, Cambridge, MA, 1969.

[4] N. N. Aizenberg, Yu. L. Ivaskiv, "Multiple-valued threshold logic". Naukova Dumka Publisher House, Kiev, 1977 (in Russian).

[5] N. N. Aizenberg, I. N. Aizenberg, and G. A. Krivosheev, "CNN based on universal binary neurons: learning algorithm with error-correction and application to impulsive-noise filtering on gray-scale images", in *the Fourth Int. Workshop on Cellular Neural Networks and their Applications*, Seville, Spain, June 24-26, 1996, pp. 309-314.

[6] T. Kim, and T. Adali, "Approximation by fully complex multilayer perceptrons", *Neural computation* 2003, vol. 15, no. 7, pp. 1641-1666.