ISSN 1064-5624, Doklady Mathematics, 2018, Vol. 98, No. 2, pp. 421–424. © Pleiades Publishing, Ltd., 2018.
Original Russian Text © A.E. Dyusembaev, M.V. Grishko, 2018, published in Doklady Akademii Nauk, 2018, Vol. 482, No. 2.

MATHEMATICS

On Correctness Conditions for Algebra of Recognition Algorithms with μ-Operators over Pattern Problems with Binary Data

A. E. Dyusembaev[a,*] **and M. V. Grishko**[a,**]

Presented by Academician of the RAS Yu.I. Zhuravlev April 6, 2018

Received April 19, 2018

Abstract—The concept of an Ω-weakly regular problem is introduced. On the basis of the Zhuravlev operator approach combined with the neural network paradigm, it is shown that, for each such problem, a correct algorithm and a six-level spatial neural network reproducing the computations executed by this algorithm can be constructed. Moreover, the set of Ω-weakly regular problems includes the set of Ω-regular problems. It turns out that a three-level spatial network (μ-block) is a forward propagation network whose inner loop under estimation of the class membership for each test object consists of a single iteration. As a result, the amount of computations required for the six-level network is reduced noticeably.

DOI: 10.1134/S1064562418060078

Conditions for the correctness of the algebra of recognition algorithms with estimate calculation μ-operators over recognition problems with binary data (referred to hereafter as problems) were determined in [3]. These conditions are sufficient ones for correctness and are formulated in the form of constraints on the set of recognition problems (Ω-regular problems) for which a correct algorithm [1] can be constructed. Moreover, each operator of the initial family (model ECO [2, 3]) is associated with a μ-block, i.e., a spatial three-level multilayered neural network [3, 4] reproducing the computations executed by an algorithm (operator) of the initial family. In this context, a common question arises as to how to relax the above-mentioned correctness conditions. Below, we introduce the concept of an Ω-weakly regular problem, which is then used, in conjunction with operator theory [1, 2] and the neural network paradigm, to give an answer to the above question. An interesting aspect associated with the obtained result is that, for the considered problems, we can construct a six-level spatial neural network such that each of its μ-blocks is a forward propagation network. Moreover, for each object of a test sample, the computational process of a μ-block (its inner loop) used to estimate class membership consists of a single iteration, which opens up an opportunity to optimize the computational process for the network as a whole. Accordingly, the goal of this study is to show the opportunity of relaxing the correctness conditions for the algebra of recognition algorithms with estimate calculation μ-operators [3] and to justify the improved efficiency of the neural network as applied to recognition problems with binary data. It is well known that these problems form a subclass in the metric space U of recognition problems [1].

Suppose that $X = \{x \mid x = (x_1, x_2, \ldots, x_n),\ x_i \in \{0,1\},\ i = 1, 2, \ldots n\}$ is the space of original objects, i.e., n-dimensional binary vectors space, and $C_1', C_2', \ldots, C_l'$ are classes entirely covering X. Let $Q_1, Q_2, \ldots, Q_\ell$ be a system of two-valued monadic predicates over X such that $Q_j(x) \equiv \langle\langle x \in C_j \rangle\rangle$, $x \in X$, $j = 1, 2, \ldots, \ell$. A recognition problem $u \in U$ is an ordered pair $u = (I_0, X^q)$, where $I_0 = \langle X^m, \alpha \rangle$ is the initial information of the problem u, $X^m = \{x^1, x^2, \ldots, x^m\}$ is a training sample of the problem u, and $\alpha = \|\alpha_{ij}\|_{m\times\ell}$ is the classification matrix of the sample $X^m(\alpha_{ij} = Q_j(x^i),\ j = 1, 2, \ldots, \ell;\ i = 1, 2, \ldots, m)$. The sample $X^q = \{\mathbf{x}^1, \mathbf{x}^2, \ldots, \mathbf{x}^q\}$ for the problem u is the sample of test objects to be recognized. The elements of the information matrix for X^q, i.e., the elements of the classification matrix $f = \|f_{ij}\|_{q\times\ell}$ of the problem u, are defined as $f_{ij} = Q_j(\mathbf{x}^i)$, $j = 1, 2, \ldots, \ell$; $i = 1, 2, \ldots, q$. Given an arbitrary problem $u = (I_0, X^q)$, assume that $X^m \cap X^q = \phi$ and $C_j = C_j' \cap X^m$, $j = 1, 2, \ldots, \ell$. As an initial equation, we use the operator equation

[a] Faculty of Mechanics and Mathematics, Al-Farabi National University of Kazakhstan, Almaty, 050040 Kazakhstan
*e-mail: anuardu@yahoo.com
**e-mail: gmv_87@mail.ru

$$\mathcal{A}u = f, \quad \mathcal{A} \in H, \quad u \in U, \quad f \in F, \tag{1}$$

which is usually associated with the solution of the direct or inverse problem, where $\mathcal{A}$ is a recognition algorithm (operator) acting from U to F. As a rule, H is specified as the parametric family of algorithms $\{\mathcal{A}_p\} = \{\mathcal{A}_p \mid p = (p_1, p_2, \ldots, p_c) \in P\}$, among which, by using a certain optimization procedure, one tries to find an algorithm guaranteeing an acceptable solution of the problem u. Since the inverse problem is usually unsolvable for Eq. (1), in the case of a normed space F, the problem of synthesizing an algorithm $\mathcal{A}$ producing the exact solution of the problem $u \in U$ is replaced by an extremal problem of the type

$$\min_{p \in P} \|\mathcal{A}_p u - f\|^2. \tag{2}$$

Its solution (if any) determines an algorithm $\mathcal{A}_{p*}$ that is an approximate solution of the inverse problem for (1). From a practical point of view, the conditions for constructing $\mathcal{A}_{p*}$ can be rather restrictive and their verification can face difficulties. In the case under consideration, using the theorem from [2] on the decomposition of the recognition operator, we replace Eq. (1) by the system

$$Au = \varphi, \quad A \in \vartheta\{A_p\}, \quad u \in U,$$

$$R^*\varphi = f, \quad \varphi \in \Phi, \quad f \in F,$$

where ϑ is the algebra over $\{A_p\}$ defined in [2]; A is the operator calculating the estimate matrix $\varphi = \|\varphi_{ij}\|_{q\times\ell}$; Φ is the metric space of $q \times \ell$ matrix; and, given the matrix φ, the decision rule R^* [2] determines to which of the classes $C_1', C_2', \ldots, C_l'$ the objects of the sample X^q belong. Thus, the problem of constructing the algorithm $\mathcal{A}^*$ is reduced to the problem of constructing $A^* \in \vartheta\{A_p\}$ such that $\mathcal{A}^* = A^* \circ R^*$ is a special case of the classical problem of operator recovery from given information on the solution and the right-hand side of an operator equation. In a number of cases [1], relying on the operator approach, the correctness conditions for the algebra ϑ over a set of recognition problems can be determined by augmenting the family $\{A_p\}$, which implies a description of the set of problems U such that, for each problem $u \in U$ in $\vartheta\{A_p\}$, there is an operator providing its exact solution, i.e., $\mathcal{A}^*u = f$, where $u \in U$, $\mathcal{A}^* = A^* \circ R^*$, and $A^* \in \vartheta\{A_p\}$. Assuming that the original set of algorithms is a family of estimate calculation algorithms [2] and u is a regular problem, an explicit expression for A^* in the form of an operator polynomial was found in [2] and $\mathcal{A}^*$ (Zhuravlev's operator) was given by

$$\mathcal{A}^* = ((\theta_1 + \theta_2)\sum_{i=1}^{q}\sum_{j=1}^{l} f_{ij} \cdot B^k(i,j)) \circ R^*(\theta_1, \theta_2) \tag{3}$$

$$k = \left[\frac{\ln q + \ln \ell + \ln(\theta_1 + \theta_2) - \ln \theta_1}{|\ln a|}\right] + 1, \tag{4}$$
$$a = \max_{i,j} \max_{(s,h)\neq(i,j)} |\Gamma_{sh}(i,j)|.$$

Equation (3) has a universal character. Note that the degree k of closure [2] can be reduced, but, in this paper, we use formula (4). In (3), given a matrix $\varphi = \|\varphi_{ij}\|_{q\times\ell}$, the threshold decision rule $R^*(\theta_1, \theta_2)$ with parameters $\theta_1 = \min \theta_{1j}$ and $\theta_2 = \max \theta_{2j}$ (in our case, with additional constraints [3, 4]) and k determined by (4) calculates the matrix $\beta = \|\beta_{ij}\|_{q\times\ell}$ coinciding with the matrix f of the problem u. In a more general case, for the recognition algorithm $\mathcal{A}$, the decision rule $R^*(\theta_{1j}, \theta_{2j})$ with parameters θ_{1j} and θ_{2j} by φ calculates a matrix $\beta = \|\beta_{ij}\|_{q\times\ell}$ that may not coincide with the matrix $f = \|f_{ij}\|_{q\times\ell}$ of the problem u. In [3] the concept of an Ω-regular recognition problem was introduced and it was shown that a correct algorithm of form (3) can be constructed for each Ω-regular problem u. The algebra ϑ [2] was constructed over the family $\mathfrak{M} = \{A(\mu, \Omega, \varepsilon, p^n, \gamma^m)\}$ of estimate calculation operators (ECO) by applying the following operations:

(a) const $\cdot A$; (b) $A' \cdot A''$; (c) $A' + A''$.

Each operator $A \in \mathfrak{M}$ was associated with a three-level spatial neural network (μ-block) that, for given $u = (I_0, X^q)$, reproduces the computation executed by A. Such an operator was called in [3] a μ-operator. The μ-block sequentially calculates the rows of the estimate matrix $\Gamma^\mu = \|\Gamma_{ij}^\mu\|_{q\times\ell}$, which is called the matrix of the μ-block, where Γ_{ij}^μ is an estimate determining the membership of the object $\mathbf{x}^i \in X^q$ in the class C_j. A number of constraints were assumed to hold, one of which was associated with the neural network paradigm. Specifically, since the synoptic strength of a biological neuron, including ones making up the μ-block, is limited, their output values have to belong to some interval, namely, the half interval $(-1, 1]$. Assume also that the elements of the operand matrices in (a)–(c) and the elements of the resulting matrices in (a)–(c) belong to the same interval.

Recall that, according to the classical approach [1, 2], the construction of a correct algorithm (see (3)) assumes the construction of quasi-basis operators $B(i, j)$, $i = 1, 2, \ldots, q$; $j = 1, 2, \ldots, \ell$. Here,

$$B(i,j) = \sum_{t\in\{1,2,\ldots,\ell\}\setminus\{j\}} B_{jt} + \sum_{\tau\in\{1,2,\ldots,\ell\}\setminus\{i\}} B_{i\tau}^j,$$

where each of the operators B_{jt} for $j = 1, 2, \ldots, \ell$ and $j = 1, 2, \ldots, \ell$; $t = 1, 2, \ldots, j-1, j+1, \ldots, \ell$ ($j = 1, t \geq 2$) is an estimate calculation μ-operator and each of the operators $B_{i\tau}^j$ for $j = 1, 2, \ldots, \ell$; $i = 1, 2, \ldots, q$; and $\tau =$

$1, 2, \ldots, i-1, i+1, \ldots, q$ ($i = \overline{1,\tau} \geq 2$) is either an estimate calculation μ-operator or the difference between two such operators.

Let U be the class of Ω-regular problems [3, 4]. Let $\Omega = \{e_a\}$ with $e_a \subset \{1, 2, \ldots, n\}$ and $e, e' \in \Omega$. Define

$$D_e = \bigcup_{e' \neq e} e'.$$

If $\Omega = e_a$, we set $D_e = \phi$. Given an object $x \in X$, a set of feature weights $p = (p_1, p_2, \ldots, p_n)$, and $e = (h_1, h_2, \ldots, h_{|e|}) \in \Omega$, we define $ex = x_e = (x_{h_1}, x_{h_2}, \ldots, x_{h_{|e|}})$ and consider the number $\rho(ex', ex'')$ of coinciding coordinates for the vectors ex', ex'', where $x', x'' \in X$. In what follows, let $n > 1$, $l > 1$, $m > 1$, and $q > 1$.

The problem $u = (I_0, X^q)$ is called Ω-weakly regular if the following conditions are satisfied:

(i) $C_{j'} \neq C_{j''} (j' \neq j'', 1 \leq j', j'' \leq \ell)$.

(ii) For each pair $\mathbf{x}, \mathbf{y} \in X^q$ $(\mathbf{x} \neq \mathbf{y})$, there are $x' \in X^m$, $\overline{e} \in \Omega$, and $i_r \in \overline{e}$ such that

(a) $i_r \notin D_{\overline{e}}$, (b) $\rho(\overline{e}x', \overline{e}\mathbf{x}) \neq \rho(\overline{e}x', \overline{e}\mathbf{y})$.

The feature $i_r \in \overline{e}$ in (ii) is called singled out feature for the pair **x**, **y** by the set Ω. The entire set of Ω-weakly regular problems is denoted by U^0. Then $U \subset U^0$. Our first goal is, relying on the operator approach [1, 2], the results of [6], and the neural network paradigm, to show and justify the possibility of constructing a correct algorithm for any problem $u \in U^0$. As original operators, we also use estimate calculation μ-operators, but the activation functions of first-level neurons of a μ-block are different from those used in [3, 4]. Let $\mathbf{x} \in X^q$, $x \in X^m$, $e = (h_1, h_2, \ldots, h_{|e|})$, and $p_e = (p_{h_1}, p_{h_2}, \ldots, p_{h_{|e|}})$. Define the constant $p_0 = \dfrac{1}{(\ell - 1) \cdot |\Omega|}$. Let the activation function of first-level neurons (as well as of their duplicates along the Z axis) in a μ-block be defined as

$$F_\mu^\varepsilon(ex, e\mathbf{x}) = \begin{cases} \mu \cdot \dfrac{(x_e, p_e)}{\sum_{s=1}^{|e|} x_{h_s}}, & \text{if} \quad \rho_e(x, \mathbf{x}) \geq n - \varepsilon \quad \text{and} \quad x_e \neq \overline{0}_e \\ \mu \cdot p_0, & \text{if} \quad \rho_e(x, \mathbf{x}) \geq n - \varepsilon \quad \text{and} \quad x_e = \overline{0}_e \\ (1-\mu) \cdot \dfrac{(x_e, p_e)}{\sum_{s=1}^{|e|} x_{h_s}}, & \text{if} \quad \rho_e(x, \mathbf{x}) < n - \varepsilon \quad \text{and} \quad x_e \neq \overline{0}_e \\ (1-\mu) \cdot p_0, & \text{if} \quad \rho_e(x, \mathbf{x}) < n - \varepsilon \quad \text{and} \quad x_e = \overline{0}_e. \end{cases}$$

Note that an activation function of this type is intended for constructing a μ-block for the operator B_{jt}. Another type of activation function used for first-level neurons of a μ-block and their duplicates along the Z axis is defined as

$$F_\mu^\varepsilon(ex, e\mathbf{x}) = \begin{cases} \mu \cdot \dfrac{(x_e, p_e)}{x_{i_r}}, & \text{if} \quad \rho_e(x, \mathbf{x}) \geq n - \varepsilon \quad \text{and} \quad x_{i_r} \neq 0 \\ \mu \cdot \dfrac{2 \cdot p_{i_r}(e)}{1 + e^{-(x_e, p_e)}}, & \text{if} \quad \rho_e(x, \mathbf{x}) \geq n - \varepsilon \quad \text{and} \quad x_{i_r} = 0 \\ (1-\mu) \cdot \dfrac{(x_e, p_e)}{x_{i_r}}, & \text{if} \quad \rho_e(x, \mathbf{x}) < n - \varepsilon \quad \text{and} \quad x_{i_r} \neq 0 \\ (1-\mu) \cdot \dfrac{2 \cdot p_{i_r}(e)}{1 + e^{-(x_e, p_e)}}, & \text{if} \quad \rho_e(x, \mathbf{x}) < n - \varepsilon \quad \text{and} \quad x_{i_r} = 0. \end{cases}$$

An activation function of this type is used to construct a μ-block for the operator $B_{i\tau}^j$. Here, i_r is the index of the singled out feature for a fixed pair of objects from the test sample. The function $p_{i_r}(e) = p_{i_r}$ if $i_r \in e$ and $p_{i_r}(e) = 0$ otherwise.

Let ϑ be the algebra of recognition algorithms constructed over μ-operators of the ECO family $\mathfrak{M} = \{A(\mu, \Omega, \varepsilon, p^n, \gamma^m)\}$ [3] by applying operations (a)–(c). The next theorems hold.

Theorem 1. *Suppose that a nonempty system $\Omega = \{e_a\}$ with $e_a \subset \{1, 2, \ldots, n\}$ is given, and let $u = (I_0, X^q) \in U^0$ be an arbitrary Ω-weakly regular problem and $f = \|f_{ij}\|_{q \times \ell}$ be the classification matrix of u. Then the algorithm $\mathcal{A}^*$*

given by (3) *with k defined by* (4) *is correct for the problem u*.

An important positive effect of this result is the opportunity of relaxing the correctness conditions and optimizing the computational process for a special class of neural networks. Here and below, we deal with the class of six-level spatial neural networks [4, 5] whose elementary blocks are μ-blocks.

Theorem 2. *Suppose that a nonempty system* $\Omega = \{e_a\}$ *with* $e_a \subset \{1, 2, \ldots, n\}$ *is given, and let* $u = (I_0, X^q) \in U^0$ *be an arbitrary* Ω*-weakly regular problem with classification matrix* $f = \|f_{ij}\|_{q\times\ell}$. *Then*, *for the problem u*, *there is a six-level spatial network with at most* $\ell \cdot (\ell - 1) + 2\ell \cdot q(q - 1)$ μ*-blocks that calculates a matrix* β *coinciding with the matrix f of the problem u*.

In what follows, let $\Gamma^\mu = \|\Gamma^\mu_{ij}\|_{q\times\ell}$ be the matrix of a μ-block and $\Omega = \{e_a\}$ with $e_a \subset \{1, 2, \ldots, n\}$ be a given nonempty system. Relying on the proof of Theorem 2 and additional information, we can formulate the following result.

Theorem 3. *Let* $u = (I_0, X^q) \in U^0$ *be an arbitrary* Ω*-weakly regular problem with classification matrix* $f = \|f_{ij}\|_{q\times\ell}$. *Then*, *for the problem u*, *a six-level spatial network with at most* $\ell \cdot (\ell - 1) + 2\ell \cdot q(q - 1)$ μ*-blocks can be constructed so that each* μ*-block is a forward propagation network whose inner loop consists of a single iteration when calculating the estimate* Γ^μ_{ij} $(i = 1, 2, \ldots, q;$ $j = 1, 2, \ldots, l)$. *Moreover*, *if the matrix* $f = \|f_{ij}\|_{q\times\ell}$ *of the problem u and the parameter k are specified*, *then the matrix* β *output by the network coincided with the matrix f.*

The proof of Theorem 1 is based on Theorem 2 from [2], while Theorem 1 itself underlies the proofs of Theorems 2 and 3. A feature of the considered networks is that their inner levels make use of diagonal activation functions, for which the output of the neuron adder coincides with the value of the activation function itself. As a result, the intermediate computations in the inner and outer loops of these networks are noticeably simplified. Note also that, as Ω, one can use $\Omega_0 = \{\{1\}, \{2\}, \ldots, \{n\}\}$ and $\Omega_1 = \{\{1, 2, \ldots, n\}\}$ [1, 2], subsets of Ω_0 of cardinality $k \geq 1$, bases of the feature set [6], representative sets [1], subsets of irredundant tests [6, 7] (for problems with disjoint classes), and so on.

Concerning the efficiency of the computational process in a network of the considered type, we note that the only delay in the computations is associated solely with the fifth level of the network, where a simple function is recursively calculated [4, 5]. The recursion does not go beyond the fifth level and does not mean recalculations of weights for neurons of this level.

REFERENCES

1. Yu. I. Zhuravlev, *Selected Papers* (Magistr, Moscow, 1998) [in Russian].
2. Yu. I. Zhuravlev, Kibernetika, No. 2, 37–45 (1978).
3. A. E. Dyusembaev and M. V. Grishko, J. Pattern Recogn. Image Anal. **27** (2), 166–174 (2017).
4. A. E. Dyusembaev and D. R. Kaliazhdarov, Dokl. Math. **91** (2), 236–239 (2015).
5. A. E. Dyusembaev, Dokl. Math. **95** (2), 125–128 (2017).
6. A. E. Dyusembaev, USSR Comput. Math. Math. Phys. **22** (6), 217–226 (1982).
7. M. A. Solov'ev, *Tests* (*Theory, Constructions, and Applications*) (Nauka, Moscow, 1978) [in Russian].

Translated by I. Ruzanova