

On Computational Complexity of Multiclass Classification Approach ECOC

Hayk, Danoyan

Institute for Informatics and Automation Problems
Yerevan, Republic of Armenia
e-mail: heddan1987@gmail.com

ABSTRACT

The multiclass classification approach known as ECOC (error correcting output codes) is considered. The method solves the multiclass classification problem by combining binary classification algorithms according to some binary matrix. The framework is considered where the columns of the mentioned matrix are selected from the given set. It is proved that the algorithmic problem of column selection optimizing the training error is NP complete.

Keywords

Computational efficiency, ECOC, Ensemble Methods, Minimum Test Collection, Multiclass Classification

1. INTRODUCTION

Classification problem is one of the major research areas in artificial intelligence, and particularly, in machine learning. Generally, classification is a procedure, which put objects into categories (classes). Usually objects belonging to the same class share common properties. So, if the class of some object is known a lot of information is available about it. Classification problem naturally arises in practical applications such as: machine vision, speech recognition, text or document classification, optical character recognition, protein function prediction, medical diagnosis, etc. This does not contain all possible applications of classification.

Assume that the objects to be classified are represented by the corresponding feature vectors belonging to a set S , which is the subset of n -dimensional vector space R^n , where R is a set of reals and n is a natural number. Assume we have a set of $l \geq 2$ classes associated to the set of labels $Y = \{1, \dots, l\}$, and let an unknown target function $f : S \rightarrow Y$ exists. It is known that function f defines a partition of set S into the l classes by the following way: $S = \cup_{i=1}^l C_i$, where $C_i = \{x \in S | f(x) = i\}$. Let us have a set of instances $C_i^* \subseteq C_i$, for each class i , $i = 1, \dots, l$ and $|C_i^*| = m_i$. It is required to find an approximation for the function f from the given set of hypothesis $H \subseteq \{h | h : S \rightarrow Y\}$, or, in other words, for a given $\tilde{x} \in S$ it is required to find the corresponding class t (such that $\tilde{x} \in C_t$).

The classification problem is well-studied for the case of two classes (binary classification). For details we refer to [1, 2]. But most of the applications require multiclass classification algorithms ($l > 2$). There are plenty

of methods for multiclass classification and theory for these methods [1, 3, 4].

We consider a multiclass to binary reduction framework known as ECOC (Error-correcting Output Codes) [5, 6] from the view of computational efficiency. Method decomposes multiclass classification problem into several binary classification problems. Each class is encoded as a binary vector of length N (N is a nonnegative integer) composed by output labels of these classifiers. We construct the matrix $\alpha_{l \times N}$ consisting of the mentioned vectors in rows. For each j th ($j = 1, \dots, N$) column of matrix $\alpha_{l \times N}$ we consider the binary classification problem with the set of instances C_0^{*j} and C_1^{*j} defined in the following way:

$$C_0^{*j} = \bigcup_{\substack{i=1 \\ \alpha_{ij}=0}}^l C_i^*, \quad C_1^{*j} = \bigcup_{\substack{i=1 \\ \alpha_{ij}=1}}^l C_i^* \quad (1)$$

So, for each column a binary classification problem is to separate an appropriate group of classes from all the others. In each column for $j = 1$ to N a binary algorithm gets as the learning set C_0^{*j} and C_1^{*j} and returns some hypothesis $h_j : S \rightarrow \{0, 1\}$. Let us denote $\beta_j = h_j(\tilde{x})$ for given $\tilde{x} \in S$. So we get the binary vector $\tilde{\beta}_{\tilde{x}} = (\beta_1, \dots, \beta_N)$ corresponding to the given object \tilde{x} . According to the considered approach the object \tilde{x} will be classified to the class t , where t is defined in the following way:

$$t = \arg \min_{i=1, \dots, l} \sum_{j=1}^N |\alpha_{ij} - \beta_j| \quad (2)$$

The object \tilde{x} is classified as the element of a class corresponding to the nearest row of matrix by means of hamming distance. For the theoretical justification of this approach we refer to [8, 9].

As it follows from (2), the classification accuracy of the considered approach depends on the choice of matrix α and the corresponding binary classification algorithms. As a good review of existing approaches we refer to [10]. A modification of boosting algorithm for ECOC was considered in [11]. For the voting algorithm the problem of constructing of good matrix was studied in [12].

The main goal of this paper is investigation of algorithmic complexity of constructing the matrix α optimizing the training error.

2. COLUMN SELECTION

The goal of this paper is to analyze the computational complexity of problem of finding the matrix α minimizing

ing the training error under assumptions that:

I. The columns of α will be selected from the given set of vectors of length l $\{\tilde{\alpha}^1, \dots, \tilde{\alpha}^M\}$, where M is natural number.

II. For each training set generated by binary vector $\tilde{\alpha}^j$ by (1) the binary classification algorithm A_j is fixed.

As it was already mentioned, the algorithm A_j gets as a training set $C_0^{*j} = \cup_{\alpha_{ij}=0}^l C_i^*$, and $C_1^{*j} = \cup_{\alpha_{ij}=1}^l C_i^*$ and returns hypothesis $h_j : S \rightarrow \{0, 1\}$. We want to construct α by selecting some vectors from the set $\{\tilde{\alpha}^1, \dots, \tilde{\alpha}^M\}$ as columns of α such that the number of correctly classified examples will be maximal. After applying binary classification algorithms on each instance \tilde{x}^{ij} , (by \tilde{x}^{ij} is denoted the j th element of the set C_i) we will transform the set of instances to the corresponding set of binary vectors of length M . Denote by $\tilde{\gamma}^{ij} = (h_1(\tilde{x}^{ij}), \dots, h_M(\tilde{x}^{ij}))$, $i = 1, \dots, l$, $j = 1, \dots, m_i$. Let us denote the rows of matrix with columns $\tilde{\alpha}^1, \dots, \tilde{\alpha}^M$ by $\tilde{c}^1, \dots, \tilde{c}^l$. For any binary vector $\tilde{z} = (z_1, \dots, z_M) \in E^M$ where $E = \{0, 1\}$ denote by $\alpha_{\tilde{z}}$ the matrix which columns are the vectors $\{\tilde{\alpha}^i / i \in \text{supp}(\tilde{z})\}$ where by $\text{supp}(\tilde{z})$ is denoted the set $\{i / i \in \{1 \dots, M\}, z_i = 1\}$. Denote by $R(\tilde{z})$ the number of correctly classified samples \tilde{x}^{ij} using matrix $\alpha_{\tilde{z}}$. It is obvious that the instance $\tilde{\gamma}^{ij}$ will be classified correctly if the following system of inequalities takes place:

$$\begin{cases} \|\tilde{z} \wedge (\tilde{\gamma}^{ij} \oplus \tilde{c}^i)\| < \|\tilde{z} \wedge (\tilde{\gamma}^{ij} \oplus \tilde{c}^k)\| \\ k = 1, \dots, l, k \neq i \end{cases} \quad (3)$$

where \oplus denotes the sum of two vectors by modulo 2, \wedge is coordinate wise logical AND, and by $\|\tilde{u}\|$ is denoted the hamming weight of vector $\tilde{u} = (u_1, \dots, u_M)$ i.e., the number of its nonzero elements. Let us define the following predicate for each $1 \leq i \leq l, 1 \leq j \leq m_i$:

$$P_{ij}(\tilde{z}) = \begin{cases} 1 & \text{if system (3) takes place} \\ 0 & \text{otherwise} \end{cases}$$

It is obvious that $R(\tilde{z}) = \sum_{i=1}^l \sum_{j=1}^{m_i} P_{ij}(\tilde{z})$, for any $\tilde{z} \in E^M$.

The condition that different classes (after column selection procedure) must be associated to different binary vectors may also be required. This may be written as:

$$\begin{cases} \|\tilde{z} \wedge (\tilde{c}^i \oplus \tilde{c}^j)\| > 0 \\ 1 \leq i < j \leq l \end{cases} \quad (4)$$

Now consider the following optimization problem: find $\tilde{\omega} \in E^M$ such that $\tilde{\omega} \neq 0^M$ (by 0^M is denoted the all zero vector of length M), system (4) takes place and

$$R(\tilde{\omega}) = \max_{\tilde{z} \in E^M} R(\tilde{z}).$$

Later we will consider the decision version of problem i.e., for the given number a it is required to find is there exists $\tilde{\omega} \neq 0^M$ such that system (4) takes place and $R(\tilde{\omega}) \geq a$? Later we will refer to this problem as CS (Column Selection) problem.

3. HARDNESS OF CS PROBLEM

The main result of this paper concerns to the algorithmic complexity of CS problem stated in Paragraph 2. We assume that the reader is acquainted with Turing machines, P/NP classes and polynomial reducibility.

We use definitions brought in [13, 14]. Let us bring a formulation of the following two NP complete problems.

Set Covering Problem [13, 15]: Let the finite set $B = \{b_1, \dots, b_u\}$, family of subsets of $\mathfrak{B} = \{B_1, \dots, B_v\}$ and nonnegative integer w be given. Are there w subsets $B_{i_1}, \dots, B_{i_w} \in \mathfrak{B}$ such that $\bigcup_{j=1}^w B_{i_j} = B$? It is known that this set covering problem is an NP complete problem [15].

Minimum Test Collection Problem [13]: Let us have a binary matrix $\mu_{p \times q}$ and nonnegative integer k . Do there exist at most k columns of $\mu_{p \times q}$ such that the matrix formed by that columns do not contain equal rows? Mention that some other statements of minimum test collection problem exist [16]. M. Garey and D. Jonson [13] claimed that in their unpublished paper it is proved that Minimum Test Collection Problem is NP complete. We were not able to find the mentioned unpublished paper and for completeness of summary we will bring the proof of the mentioned fact.

Theorem 1. Minimum Test Collection Problem is NP complete.

Proof: As it is known the proof of NP completeness consists of two steps. The first step is to prove that the problem belongs to class NP, and the second is the polynomial reduction of a known NP complete problem to problem under consideration [13, 14, 15]. It is easy to see that the Minimum test collection problem belongs to the class NP. Now let the particular instance of set covering problem (later related as SCI) be given, i.e., we have the set $B = \{b_1, \dots, b_u\}$, family of subsets of $\mathfrak{B} = \{B_1, \dots, B_v\}$ and nonnegative integer w . It is required to check if there exists a family of subsets B_{i_1}, \dots, B_{i_w} such that $\bigcup_{j=1}^w B_{i_j} = B$. Let us consider the following particular instance of the minimum test collection problem (MTCI) : the matrix $\mu_{p \times q}$ for MTCI with $p = 2u$, and $q = v + u - 1$ is defined as it shown in Figure 1.

| | 1 | ... | v | v+1 | ... | v+u-1 |
|-----|-----------------------|-----|---|-----------|-----|-------|
| 1 | | | | | | |
| ⋮ | $\delta_{u \times v}$ | | | I_{u-1} | | |
| u | | | | 0^{u-1} | | |
| u+1 | | | | | | |
| ⋮ | $0_{u \times v}$ | | | I_{u-1} | | |
| 2u | | | | 0^{u-1} | | |

Figure 1: instance of MTC problem corresponding to instance of SC problem.

The matrix $\delta_{u \times v}$ is defined in the following way:

$$\delta_{ij} = \begin{cases} 1, & \text{if } b_i \in B_j \\ 0, & \text{otherwise,} \end{cases}$$

for $1 \leq i \leq u, 1 \leq j \leq v$. By $0_{u \times v}$ is denoted the matrix of size $u \times v$ containing all 0s and by I_{u-1} we denote the identity matrix of size $(u-1) \times (v-1)$ containing ones on main diagonal and zeros, otherwise. By $\tilde{0}^{u-1}$ we denote the all zero vector of length $u-1$. We will take $k = w + u - 1$ in MTCI.

The theorem will be proved if we show the following fact: for the instance SCI there exists some w subsets B_{i_1}, \dots, B_{i_w} such that $\bigcup_{j=1}^w B_{i_j} = B$ if and only if for the MTCI there exists at most $w + u - 1$ columns of $\mu_{2u \times (v+u-1)}$ such that the matrix formed by those rows does not contain the equal rows.

Assume that we are given $B_{i_1}, \dots, B_{i_w} \in \mathfrak{B}$ such that $\bigcup_{j=1}^w B_{i_j} = B$ then by definition of matrix $\delta_{u \times v}$ we get that the matrix composed of columns by numbers i_1, \dots, i_w of $\delta_{u \times v}$ does not contain all zero vector. Now consider the matrix Δ composed of columns of μ by numbers $i_1, \dots, i_w, v+1, \dots, v+u-1$. From definition we get that the rows of Δ from 1st to u th differ one from the other and the rows from $u+1$ to $2u$ also differ one from another. The t^{th} ($1 \leq t \leq u$) row of Δ may only coincide with the $(t+u)^{\text{th}}$ row but as $\delta_{u \times v}$ does not contain all zero vector, the mentioned case has also been excluded. So, we get that all rows of Δ differ one from the other.

Now let the matrix μ contain s , ($s \leq w + u - 1$) columns such that the matrix composed of those columns does not contain equal rows. It is easy to see that columns by numbers $v+1, \dots, v+u-1$ must necessarily be contained in that set of s columns. Let those columns be $i_1, \dots, i_r, v, \dots, v+u-1$. As the obtained matrix does not contain equal rows then from the definition of μ follows that the matrix composed of columns by numbers i_1, \dots, i_r of δ does not contain the all zero vector (because, otherwise, if the t th row of mentioned matrix is the all zero vector then the t th row of the matrix composed by columns by numbers $i_1, \dots, i_r, v, \dots, v+u-1$ of μ will be equal to the $(u+t)^{\text{th}}$ row which is a contradiction). This means that $\bigcup_{j=1}^r B_{i_j} = B$ where $r \leq w$. Then there exist w subsets belonging to \mathfrak{B} covering the set B . The theorem is proved.

Now we can formulate the main result of this paper:

Theorem 2. The CS problem is NP complete.

Proof. Again, the theorem will be proved if we show that the problem belongs to the class NP and every particular instance of some known NP complete problem may be reduced in polynomial time to the certain instance of the problem CS [13, 15]. The fact that the problem MB belongs to the class NP is obvious. Let us consider the particular instance of Minimum Test Collection, which is an NP complete problem by the Theorem 1. Let we have a binary matrix $\mu_{p \times q}$ and nonnegative integer k . One wants to know if there are at most k columns such that each pair of rows of matrix formed by those columns of μ differs at least in one coordinate. We will prove the theorem by reducing the mentioned problem to some instance of the problem CS. Later we

will refer to the mentioned instance as CSI. CSI is defined in the following way (see Figure 2):

$M = 3q + k + 1, l = 2q + p + k + 3, \tilde{c}^i = (\tilde{\mu}^i | \tilde{1}^{k+1+2q})$, $\tilde{\mu}^i = (\mu_{i1}, \dots, \mu_{iq})$ $i = 1, \dots, p$. Then, $\tilde{c}^{p+i} = (\tilde{\mu}^i | \tilde{\lambda}^i)$, where $i \in \{1, \dots, 2q + k + 1\}$ and $\tilde{\lambda}^i \in E^{2q+k+1}$ is a vector of all ones besides the i th coordinate; $\tilde{c}^{l-1} = (\tilde{1}^{q+k+1} | \tilde{0}^{2q})$, $\tilde{c}^l = \tilde{0}^{3q+k+1}$. Let $m_1 = \dots = m_l = 1$, then we will use the notation $\tilde{\gamma}^i$ instead of $\tilde{\gamma}^{i1}$ and $\tilde{\gamma}^i = \tilde{c}^i, i = 1, \dots, l-1, \gamma^l = (\tilde{1}^q | \tilde{0}^{k+1+2q})$.

| | 1 | ... | q | q+1 | ... | q+k+1 | q+k+2 | ... | 3q+k+1 |
|-------------------|-------------------------------|-----|---|--------------------------------|-----|---------------------------|-------|-----|--------|
| \tilde{c}^1 | | | | | | | | | |
| \tilde{c}^2 | μ | | | $\mathbf{1}_{p \times 2q+k+1}$ | | | | | |
| \vdots | | | | | | | | | |
| \tilde{c}^p | | | | | | | | | |
| \tilde{c}^{p+1} | $(\mu_{11}, \dots, \mu_{1q})$ | | | J_{2q+k+1} | | | | | |
| \tilde{c}^{p+2} | $(\mu_{11}, \dots, \mu_{1q})$ | | | | | | | | |
| \vdots | \vdots | | | | | | | | |
| \tilde{c}^{l-2} | $(\mu_{11}, \dots, \mu_{1q})$ | | | | | | | | |
| \tilde{c}^{l-1} | $\tilde{1}^q$ | | | $\tilde{1}^{k+1}$ | | $\tilde{0}_{2 \times 2q}$ | | | |
| \tilde{c}^l | $\tilde{0}^q$ | | | $\tilde{0}^{k+1}$ | | | | | |

Figure 2: instance of CS problem corresponding to instance of MSC problem.

In Figure 2 by J_t was denoted the matrix of size $t \times t$ containing zeros on the main diagonal and ones otherwise. It is required to decide if there exists $\omega = (\omega_1, \dots, \omega_{3q+k+1}) \in E^{3q+k+1}$ such that $R(\omega_1, \dots, \omega_{3q+k+1}) \geq l$. If $R(\omega_1, \dots, \omega_{3q+k+1}) \geq l$ then system (4) takes place for CSI.

It is easy to see that the theorem will be proved if we show the following two conditions are equivalent:

- For the MTCI there exists at most k columns such that the matrix composed of those columns does not contain equal rows.
- For CSI there exists $\tilde{\omega} = (\omega_1, \dots, \omega_{3q+k+1}) \in E^{3q+k+1}, \tilde{\omega} \neq 0^{3q+k+1}$ such that $R(\omega_1, \dots, \omega_{3q+k+1}) \geq l$.

Condition (a) means that there exists some $\tilde{\omega} = (\omega_1, \dots, \omega_{3q+k+1}) \in E^{3q+k+1}$ such that the following system takes place

$$\begin{cases} z_1 + \dots + z_q \leq k \\ \|(z_1, \dots, z_q) \wedge (\tilde{\mu}^i \oplus \tilde{\mu}^j)\| > 0 \\ 1 \leq i < j \leq p \end{cases} \quad (5)$$

when substituting $z_1 = \omega_1, \dots, z_{3q+k+1} = \omega_{3q+k+1}$.

Now assume condition a takes place i.e., there exists t ($t \leq k$) columns such that the matrix composed of those columns does not contain equal rows. Let the numbers of that columns be s_1, \dots, s_t . Consider the

vector $(\omega_1, \dots, \omega_{3q+k+1})$ defined as

$$\omega_j = \begin{cases} 1, & \text{if } j \in \{s_1, \dots, s_t\} \\ 0, & \text{otherwise,} \end{cases}$$

for $j = 1, \dots, q$ and $w_{q+1} = \dots = w_{3q+k+1} = 1$.

Now for each $i = 1, \dots, l$ consider the system

$$\begin{cases} \|\tilde{z} \wedge (\tilde{\gamma}^i \oplus \tilde{c}^i)\| < \|\tilde{z} \wedge (\tilde{\gamma}^i \oplus \tilde{c}^j)\| \\ j = 1, \dots, l, j \neq i \end{cases} \quad (6)$$

Inequality $R(\tilde{\omega}) \geq l$ is equivalent that system (6) for $\tilde{\omega}$ should take place for all $i = 1, \dots, l$. As $\tilde{\gamma}^i = \tilde{c}^i$ for $i = 1, \dots, l-1$ we can write system (6) as:

$$\begin{cases} 0 < \|\tilde{z} \wedge (\tilde{c}^i \oplus \tilde{c}^j)\| \\ j = 1, \dots, l, j \neq i \end{cases} \quad (7)$$

For $i = l$ as $\tilde{c}^l = \tilde{0}^M$ and $\tilde{\gamma}^l = (\tilde{1}^q | \tilde{0}^{k+1+3q})$ system (6) may be written as:

$$\begin{cases} z_1 + \dots + z_q < \|\tilde{z} \wedge (\tilde{1}^q | \tilde{0}^{k+1+2q} \oplus \tilde{c}^j)\| \\ j = 1, \dots, l-2 \\ z_1 + \dots + z_q < \|\tilde{z} \wedge (\tilde{\gamma}^l \oplus \tilde{c}^{l-1})\| \end{cases} \quad (8)$$

It is easy to check that vector $\tilde{\omega}$ satisfies system (7) for $i = 1, \dots, l-1$ and system (8). So, from condition a the condition b follows.

Now let the condition b takes place i.e., there exists $\tilde{\omega} \in E^{3q+k+1}$ such that $R(\tilde{\omega}) \geq l$. This means that $\tilde{\omega}$ satisfies the system (6) for all $i = 1, \dots, l$. Now let us consider the system (6) for case $i = 1$:

$$\begin{cases} 0 < \|\tilde{z} \wedge (\tilde{c}^1 \oplus \tilde{c}^j)\| \\ j = 1, \dots, l, j \neq 1 \end{cases}$$

when $j = p+t$ for $t = 1, \dots, 2q+k+1$ from Figure 2 we have that $\|\tilde{c}^1 \oplus \tilde{c}^j\| = 1$ and vectors \tilde{c}^1 and \tilde{c}^j differs only at j th coordinate so the considered inequality is $0 < \omega_j$ for $j = q+1, \dots, 3q+k+1$. Now let us write for $\tilde{\omega}$ the $(l-1)^{th}$ inequality of the system (6) for $i = l$:

$$\|\tilde{\omega} \wedge (\tilde{\gamma}^l \oplus \tilde{c}^l)\| < \|\tilde{\omega} \wedge (\tilde{\gamma}^l \oplus \tilde{c}^{l-1})\| \quad (9)$$

By Figure 2 $\tilde{\gamma}^l \oplus \tilde{c}^l = (\tilde{1}^q | \tilde{0}^{2q+k+1})$ and $\tilde{\gamma}^l \oplus \tilde{c}^{l-1} = (\tilde{0}^q | \tilde{1}^{k+1} | \tilde{0}^{2q})$, so, (9) may be written as:

$$\omega_1 + \dots + \omega_q < \omega_{q+1} + \dots + \omega_{q+k+1} \quad (10)$$

And as we already mentioned if (6) takes place for $i = 1$ then $\omega_{q+1} = \dots = \omega_{3q+k+1} = 1$. So, from (10) it follows that $\omega_1 + \dots + \omega_q \leq k$. Now let us write the first $q-1$ inequalities of system (6) for $i = 1, \dots, p$:

$$\begin{cases} \|\tilde{z} \wedge (\tilde{\gamma}^i \oplus \tilde{c}^i)\| < \|\tilde{z} \wedge (\tilde{\gamma}^i \oplus \tilde{c}^j)\| \\ j = 1, \dots, p, j \neq i \end{cases} \quad (11)$$

As $\tilde{\gamma}^i = \tilde{c}^i$, $\tilde{c}^i = (\tilde{\mu}^i | \tilde{1}^{k+1+2q})$ from (11) it follows that:

$$\begin{cases} \|(\omega_1, \dots, \omega_q) \wedge (\tilde{\mu}^i \oplus \tilde{\mu}^j)\| > 0 \\ 1 \leq i < j \leq p \end{cases}$$

so we get that the following takes place for $\tilde{\omega}$:

$$\begin{cases} \omega_1 + \dots + \omega_q \leq k \\ \|(\omega_1, \dots, \omega_q) \wedge (\tilde{\mu}^i \oplus \tilde{\mu}^j)\| > 0 \\ 1 \leq i < j \leq p, \end{cases}$$

which means that from condition b the condition a follows. The theorem is proved.

4. ACKNOWLEDGEMENT

The author would like to acknowledge the support for this project from the State committee of science, Ministry of Education and Science of the Republic of Armenia under Grant 15RF-083.

REFERENCES

- [1] A. M. Mohri, A. Rostamizadeh and A. Talwalkar, Foundations of Machine Learning. MIT press, Cambridge, USA, 432 pages, 2012
- [2] Yu. Zhuravlev, Selected scientific works, Magister, Moscow, 418 pages, 1998, in Russian
- [3] A. Daniely, S. Sabato, S. Shalev-Shwartz, Multiclass Learning Approaches: A Theoretical Comparison with Implications, in *Advances in Neural Information Processing Systems 25 (NIPS 2012)*, pp. 485-493, 2012,
- [4] B. Natarajan, Machine Learning: A Theoretical Approach, M. Kaufmann Publishers, 217 pages, 1991
- [5] T. Dietterich, G. Bakiri, Solving multiclass learning problems via error-correcting output codes, *Journal of Artificial Intelligence Research*, vol. 2. pp. 263-286, 1995
- [6] T. Sejnowski and C. Rosenberg, Parallel networks that learn to pronounce English text, *Journal of complex systems*, vol. 1, issue 1, pp. 145-168, 1987
- [7] T. Dietterich, Machine Learning Research: Four Current Directions, *AI Magazine*, vol. 18, issue 4, pp. 97-136, 1997
- [8] E. Allwein, R. Schapire, Y. Singer, Reducing Multiclass to Binary: A Unifying Approach for Margin Classifiers, *Journal of Machine Learning Research*, vol.1, pp. 113-141, 2000
- [9] E. Kong, T. Dietterich, Error-Correcting Output Coding Corrects Bias and Variance, *In Proceedings of the Twelfth International Conference on Machine Learning*, pp. 313-321, 1995
- [10] S. Escalera, O. Pujol, P. Radeva, Error-Correcting Output Codes Library, *Journal of Machine Learning Research*, no. 11, pp. 661-664, 2010
- [11] R. Schapire, Using Output Codes to boost Multiclass Learning Problems, *Proceedings of ICML-97*, pp. 313-321, 1997
- [12] V. Ryazanov, Optimization of multiclass supervised classification based on using output codes with error correcting, *Pattern Recognition and Image Analysis*, vol. 26, issue 2, pp. 262-265, 2016
- [13] M. Garey, D. Jonson, Computers and Intractability A Guide to the Theory of NP Completeness, W. H. Freeman and company, 1st edition, New York, 340 pages, 1979
- [14] S. Arora, B. Barak, Computational Complexity A Modern Approach, Cambridge University Press, 1st edition, New York, 608 pages, 2009
- [15] R. Karp, Reducibility Among Combinatorial Problems, *Complexity of Computer Computations*, pp. 85-103, 1972
- [16] I.A. Cheguis, S.V. Yablonskii, Logical methods of control of work of electric schemes, *Trudy Mat. Inst. Steklov.* Vol. 51, pp. 270-360, 1958, in Russian