Estimation of Informativeness of Recognition Signs at Extreme Information Machine Learning of Knowledge Control System

Anatoliy Dovbysh [0000-0003-1829-3318], Ihor Shelehov [0000-0003-4304-7768], Svitlana Pylypenko [0000-0003-0978-9602], Oleh Berest [0000-0003-0183-9871]

Sumy State University, Sumy, Ukraine
e-mail:s.pylypenko@media.sumdu.edu.ua

Abstract. The method of evaluating the informativeness of recognition signs is considered within the framework of information-extreme intellectual technology of machine learning. Evaluation for a student's response to a test task is accepted as a sign of recognition. Optimization of signs vocabulary is carried out directly in the process of machine learning of the knowledge control system. As a criterion for optimization of machine learning parameters, the modified Kulbak information measure is considered. Based on the optimal in the informational sense geometric parameters, that were received in the process of machine learning, of embedded hyperspherical containers of recognition classes, the unimodal deciding rules have been constructed. Verification of the effectiveness of the proposed method was carried out for the task of assessing the informativeness of tests on machine knowledge control of a particular discipline.

Keywords: extreme information intellectual technology, knowledge control system, machine learning, information criterion, unimodal classifier, vocabulary of recognition signs, informativeness of recognition sings.

1 Introduction

Wide introduction of electronic means of machine knowledge control has conditioned the relevance of the solution to the problem of adequacy of evaluation of the actual level of knowledge of the person being examined [1]. One of the key tasks aimed at solving this problem is assessment of tests informativity [2]. With the development of computerized systems in education, intelligent information technologies that can simulate the cognitive processes inherent in a person when making decision classifications became widely used when solving this task [3]. At the same time, the methods of artificial neural networks have become the most widespread [4] the content of the discipline. The analysis of existing data mining methods shows that they are mostly of a model nature and do not sufficiently take into account the abovementioned conditions of the real educational process.

One of the promising directions of increasing the functional efficiency of machine knowledge control systems is application of ideas and methods of the so-called extreme information intellectual technology (IEI-technology) data analysis, which is based on maximizing the information capacity of the system in the process of machine learning [5, 6].

The article considers the method of information-extreme machine learning of computerized knowledge control system with optimization (here and further in the informational sense) of geometric parameters of hyperspherical containers of classes of recognition with the built-in structure.

2 Problem statement

Let's consider in the framework of IEI-technology the formalized statement of the problem of information synthesis of adaptive knowledge control system with optimization of the recognition signs vocabulary. Evaluation of the response to a test, which is calculated by the evaluation function on the scale of 0 to 100, is taken for a sign of recognition. Let's imagine we have the character set $\{X_m^o \mid m=\overline{1,M}\}$ of classes of recognition, which characterize different levels of student knowledge, and the three-dimensional learning matrix of student responses to test assignments $\| y_{m,i}^{(j)} \|, \; i=\overline{1,N}, \; j=\overline{1,n}$, are given, where N is number of recognition signs, which is determined by the number of tasks in the test; n is the number of vector-implementations of classes of recognition, which ensures the representativeness of the training sample. With that, the line of the learning matrix $\{y_{m,i}^{(j)} \mid i=\overline{1,N}\}$ defines the j vector-implementation of the recognition class, and column $\{y_{m,i}^{(j)} \mid j=\overline{1,n}\}$ determines a training sample of the values of the i sign. In addition, the structured vector of machine learning parameters of the knowledge control system is known

$$g_m = < R1_m, R2_m, \delta, \Sigma^{|N|} >, \; m=\overline{1,M}, \qquad (1)$$

where $R1_m$ is inner radius of container of class X_m^o; $R2_m$ is external radius of container of class X_m^o; δ is a parameter that determines the lower control attribute for recognition signs; $\Sigma^{|N|}$ in the informational sense, is redundant initial vocabulary of power N.

In the process of machine learning of the knowledge control system, it is necessary to determine the optimal values of the coordinates of the vector (1), which provide the maximum average of the character set recognition classes of the information criterion for learning parameters optimization.

$$\overline{E}^* = \frac{1}{M} \sum_{m=1}^{M} \max_{G_E \cap \{k\}} E_m^{(k)} , \qquad (2)$$

where $E_m^{(k)}$ is information criterion, calculated on the k -th step of machine learning, for optimizing the parameters of the knowledge control system functioning when recognizing implementation of the class $X_m^o; G_E$ is working (admissible) area of determining the function of the information criterion; $\{k\}$ is an ordered set of steps in machine learning. With functioning knowledge control system in the examination mode, that is, directly in the machine knowledge control according to test tasks, it is necessary to determine the affiliation of the vector-implementation, developed-by-optimal-dictionary, to one of the recognition classes from given character set $\{X_m^o\}$.

3 Mathematical model

Since the process of knowledge evaluation is poorly formalized, the mathematical model of machine learning of the knowledge control system with the optimization of the vocabulary of recognition signs will be presented as a categorical model. In this case, the structure of the input mathematical description of the knowledge control system has the form

$$\Delta_B =< G, T, \Omega, Z, Y, X; f_1, f_2 > ,$$

where G is input factors space; T is a set of moments of information removal time; Ω is recognition signs space; Z is space of given levels of knowledge; Y is sampling population – input learning matrix; X is binary learning matrix; f_1 – is operator of the input learning matrix Y formation; f_2 is is operator of forming a working binary learning matrix X .

Fig. 1 in the form of a directed graph, shows a categorical model of machine learning of the knowledge control system with optimization of the recognition signs of vocabulary.

$$\begin{tikzcd}
\Sigma \arrow[dd, "\sigma_2"'] & & & & \\
& D \arrow[d, "\delta_2"'] & & E \arrow[lll, "\sigma_1"', bend right] \arrow[ll, "\delta_1"'] \arrow[d, "r_1"'] \arrow[dlll, bend right, "u"'] & \mathfrak{I}^{|q|} \arrow[l, "\varphi"'] \\
G \times T \times Z \times \Omega \arrow[r, "f_1"] & Y \arrow[r, "f_2"] & X \arrow[r, "\theta"] & \tilde{\mathfrak{R}}^{|M|} \arrow[r, "\psi"] & I^{|l|} \arrow[u, "\gamma"']
\end{tikzcd}$$

Fig. 1. Categorical model of machine learning with optimization of recognition signs of vocabulary.

In Fig. 1, Cartesian product $G \times T \times \Omega \times Z$ sets the universe of tests. Operator θ reflects the vectors-implementations of the binary learning matrix X partitioning $\tilde{\mathfrak{R}}^{|M|}$ of the space of signs into recognition classes, which is generally fuzzy. Operator ψ checks the main statistical hypothesis about affiliation of the vector-implementation $x_m^{(j)}$ to the corresponding recognition class. As a result of statistical verification, a set of hypotheses is formed $I^{|L|}$, where L is the number of statistical hypotheses, and operator γ forms a set of precise characteristics $\Im^{|Q|}$ where $Q = L^2$. Operator φ calculates the value E of the information criterion for optimization of the learning parameters, and operator r reconstructs the recognition class containers that are built in the radial basis of the signs space at each step of machine learning. Category model, shown in Fig. 1, has additional contours for optimizing machine learning parameters (1). The circuit, which is closed through a term-set D, a system of control tolerances, optimizes control tolerances for recognition signs. The following contour optimizes the vocabulary Σ of recognition signs. Operator $\sigma_1 : E \to \Sigma$ removes recognition signs from the vocabulary, creates new vocabularies and determines the informative nature of the recognition signs. In this case, a sign is considered informative if the value of the information criterion decreases at its removal. If the value of the criterion does not change, then the sign is considered non-informative and after the completion of machine learning the system is removed from the vocabulary. With the help of the operator $\sigma_2 : \Sigma \to \Omega$ the dimensionality of the recognition signs varies and, accordingly, a new learning matrix Y is formed. Operator u regulates the process of machine learning of the knowledge control system.

Thus, the proposed categorical model is essentially a generalized structural scheme of the information-extreme machine learning of the knowledge control system with optimization of a set of test tasks.

4 Machine learning algorithm

According to the categorical model (Fig. 1), information-extreme machine learning with optimization of the vocabulary of recognition signs is carried out according to the tricyclic procedure for finding the global maximum of the information criterion (3) for optimizing the machine learning parameters (3) in the form

$$\Sigma^* = \arg \max_{\{k\}} \{ \max_{G_\delta} \{ \max_{G_{\overline{E}} \cap \{d\}} \overline{E}^{(k)}(d) \} \}, \tag{3}$$

where $\overline{E}^{(k)}(d)$ is the averaged by the character set of recognition classes of the information criterion value (2), calculated on the a k-th step of machine learning; G_δ is the permissible area of determination of the parameter δ of the control tolerance fields for the recognition signs; $\{d\}$ is a set of values of the radii of the recognition classes containers.

The internal cycle of the algorithm (3) implements the basic algorithm of machine learning, which optimizes the geometric parameters of the recognition classes containers. For an enclosed hyperspherical container, for instance, the recognition class X_m^o the optimization parameters are the internal radius $R1_m$, closest to the center of the scattering of the recognition class vectors, and the outer radius $R2_m$. The input data for the basic machine learning algorithm is the three-dimensional array of the input learning matrix $\{y_{m,i}^{(j)} \mid m=\overline{1,M}\}; i=\overline{1,N}; j=\overline{1,n}\},$ recognition signs of which form the initial vocabulary $\Sigma^{|N|}$. Besides, the N is dimensional unit vector x_0 is given, the vertex of which determines the center of scattering of vectors-implementations of all recognition classes.

Let's consider the main stages of implementation of the basic machine learning algorithm:

1. Formation of an array of binary learning matrix $\{x_{m,i}^{(j)}\}$, the elements of which are determined by the rule: $x_{m,i}^{(j)} = \begin{cases} 1, & if \quad A_{HK,i} \le y_{m,i}^{(j)} \le A_{BK,i}; \\ 0, & if \;\; else. \end{cases}$
2. Initialization of the counter of recognition classes: $m=0$;
3. $m=m+1$;
4. Initialization of the counter of the external radius optimization steps $R2_m = 0$;
5. $R2_m = R2_m + 1$;
6. Computing the information criterion (3) and finding its maximum value E_m^* the working (admissible) function domain;
7. If $(E_m < E_m^*) \& (R2_m < N)$, then see point 5, otherwise - point 8;
8. If $E_m^* = \max\limits_{G_E \cap G_{R2}} E_m(R2),$ then $R2_m$ assign the optimal value $R2_m^*$;
9. If $m \le M$, then see point 3, otherwise - point 10, since the external radii of the containers of all classes of recognition are determined;
10. Initialization of the class of recognition counter: $m=0$;
11. $m=m+1$;
12. Initialization of the counter of internal radius optimization steps: $R1_m = 0$;
13. If $m=1,$ then $R1_1^* = 0,$ otherwise point 14 is implemented;
14. $R1_m = R1_m + 1$;
15. Calculation and search for the maximum value of the information criterion (3);
16. If $(E_m < E_m^*) \& (R1_m < R2_m)$, then see point 14, otherwise - point 17;
17. If $E_m^* = \max\limits_{G_E \cap G_{R1}} E_m(R1),$ then $R1_m$ assign the optimal value $R1_m^*$;
18. If $m \le M$, then point 11 is implemented, otherwise point 19;
19. STOP.

As a criterion for optimizing machine learning parameters in IEI-technology, Shannon entropy measure and the Kulbak information measure are mainly used, the modification of which is as follows [6]:

$$E_m^{(k)} = 0{,}5\log_2\left(\frac{D_1^{(k)} + D_2^{(k)} + 10^{-r}}{\alpha^{(k)} + \beta^{(k)} + 10^{-r}}\right)\left[\left(D_1^{(k)} + D_2^{(k)}\right) - \left(\alpha^{(k)} + \beta^{(k)}\right)\right], \qquad (4)$$

where $D_1^{(k)}$ is the first and second authenticity of classification decisions, calculated at k -th step of machine learning; $D_2^{(k)}$ is second authenticity; $\alpha^{(k)}$ is error of the first kind; $\beta^{(k)}$ is error of the second kind; 10^{-r} is a sufficiently small number to avoid division into zero (in practice, the value of r s selected from the interval $1 < r \leq 3)$.

Thus, the basic function of the basic learning algorithm in the framework of the IEI-technology is computation of the information criterion at each step of learning, organization of finding a global maximum in the work area defining the function of the criterion and determining the optimal geometric parameters of the classes of recognition containers. To optimize the recognition vocabulary, let's apply a classical scheme of sequential hereditary selection. Let's consider the main stages of the algorithm implementation for evaluating the informativeness of recognition signs in the process of information-extreme machine learning by parallel optimization of control tolerances on recognition signs:

1. Initialization of the counter of changing steps of the control tolerances field parameter on the recognition signs: $\delta = 0$;
2. $\delta = \delta + 1$;
3. Lower A_{HK} control attestation for recognition of recognition: $A_{HK} = A_{BK} - \delta$, where A_{BK} is upper control tolerance equal to the maximum value of the rating scale.
4. Initialization of the counter of recognition signs which are deleted from the vocabulary in the current step and returned again in the next step: $i = 0$;
5. The basic algorithm is implemented according to the above mentioned scheme and the averaged maximum value of the information criterion (5) is calculated by the character set of the recognition classes for the initial signs vocabulary;
6. If $\overline{E}_i \leq \overline{E}_i^*$, then point 2 is implemented, otherwise point 7;
7. $i = i + 1$;
8. If $i \leq N$, then point 5 is implemented, otherwise point 9. Informative recognition signs for which the condition is fulfilled, are identified

$$\Delta P_i = \overline{E}_{ALL} - \overline{E_i^*} > 0 \qquad (5)$$

where $\overline{E}_{ALL}$ is the maximum average value of the information criterion for optimizing machine learning parameters for the initial signs recognition vocabulary;

9. STOP.

Based on the optimal geometric parameters of embedded hyperspherical containers of recognition classes, received in the process of machine learning, unimodal deciding rules, which, unlike polymodal, are characterized by a single center of scattering of vectors-implementations of all recognition classes, are built. In this case, the deciding rules for determining the implementation of $x^{(j)}$, in the mode of examination, for example, class X_m^o have the form

$$(\forall x^{(j)} \in \tilde{\Re}^{|M|})(\forall X_m^o \in \tilde{\Re}^{|M|})\{if\ R1_m < d[x_0 \oplus x^{(j)}] < R2_m$$

$$then\ \ x^{(j)} \in X_m^o\ \ else\ x^{(j)} \notin X_m^o\},$$

where $d[x_0 \oplus x^{(j)}]$ is code distance between implementation $x^{(j)}$ and a single vector-implementation x_0, the vertex of which defines the general center of scattering of vectors-implementations of all recognition classes. Thus, information-extreme machine learning consists in purposeful approximation of the information criterion for optimizing the parameters of the system's operation to its maximum limit value.

5 Simulation results

Implementation of the above algorithm of machine learning was carried out on the example of estimating the informativeness of the tests in the discipline "Intelligent Systems", which is offered to students of Sumy State University (Sumy, Ukraine) in the specialty "Computer Science". The character set consisted of four classes of recognition: class X_1^o characterized the level of knowledge "excellent", class X_2^o – “good”, class X_3^o – “satisfactory” and class X_4^o – “poor”. The answer for each of the tests was evaluated on a 100 point scale. The complete initial vocabulary consisted of 28 recognition signs, which were equal to the number of tasks in the test. The number of vector-implementations for each of the recognition classes was 40, that is, to form the input educational matrix. The maximum average value of the information criterion calculated in the process of machine learning according to formula (4) was used to evaluate the informativeness of the tests.

In Fig. 2, shaded sections of the graph denote the work areas in which the search for a global maximum of the information criterion for optimizing machine learning parameters is is carried out. Analysis of Fig. 2 shows that the optimal value of the parameter of the control tolerances field is equal to $\delta^* = 36$ gradations of the 100-point grading scale at the maximum value of the averaged criterion. $\overline{E}^* = 1{,}35$. With the optimal parameter of the control tolerances fields on the recognition signs in accordance with procedure (5), the following optimal radii of the recognition class containers were obtained: for container recognition class X_1^o («excelent») internal radius was equal to $R1_1 = 0$ and an external radius – $R2_1 = 13$ (here and further in code units from the geometric center of the recognition class), for container recognition class

X_2^o («good») the radii were equal respectively $R1_2 = 12$ and $R2_2 = 15$, for container recognition class X_3^o («satisfactory») – $R1_3 = 15$ and $R2_3 = 17$, for container recognition class X_4^o («poor») – $R1_4 = 17$ and $R2_4 = 22$.

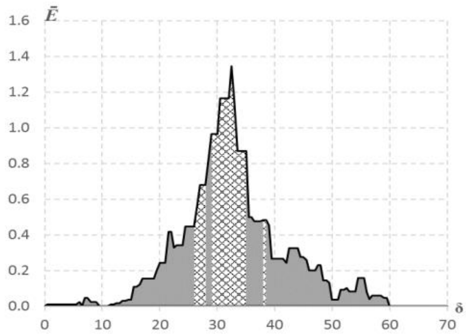

Fig. 2. The graph of the information criterion dependence on the parameter of the control tolerances field on the signs of recognition.

Fig. 2 shows the graph of dependence of the averaged information criterion (7) on the parameter δ of the control tolerance field on the recognition signs which is obtained during the machine learning process.

The optimum parameters of machine learning corresponded to the following maximum values of the information criterion (7) and the exact characteristics: for the recognition class X_1^o the maximum value of the information criterion (6) was equal to $E_1^* = 1{,}93$ at the first authenticity $D_{1,1}^* = 0{,}95$ and error of the second kind $\beta_1^* = 0{,}23$, for class X_2^o is $E_2^* = 1{,}18$ ($D_{1,2}^* = 0{,}72$, $\beta_2^* = 0{,}13$), for class X_3^o is $E_3^* = 0{,}87$ ($D_{1,3}^* = 0{,}78$, $\beta_3^* = 0{,}25$) and for class X_4^o is $E_4^* = 1{,}43$ ($D_{1,4}^* = 0{,}80$, $\beta_4^* = 0{,}15$). According to the algorithm (3), the evaluation of the informative nature of the tasks in the test was carried out by forming at each step of the machine learning new versions of vocabularies for which the optimization of the learning parameters was conducted and the maximum average value of the information criterion was carried out.

Fig. 3 shows the dynamics of the change of the averaged information criterion (6) in the process of evaluating the informativeness of the recognition signs.

Analysis of Fig. 3 shows that for the knowledge control system on the relevant discipline of twenty eight signs of information recognition, there is twenty one sign, which are indicated by light grey color. Sign 1 is non-informative because its deletion does not change the information criterion. Other signs (9, 11, 19, 22, 26 and 28) refer to interfering (disinfecting) ones, indicating the need for the processing of appropriate tasks in the test or their removal.

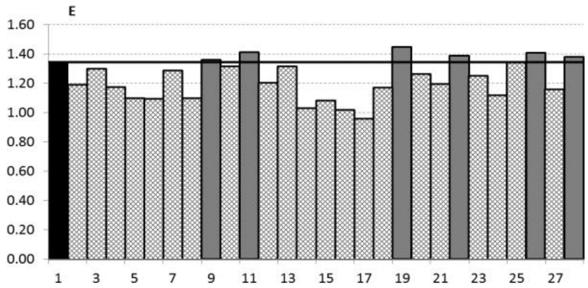

Fig. 3. Diagram of changing the information criterion in machine learning.

The most informative is the sign 17. Deleting this sign from the vocabulary leads not only to a decrease in the maximum value of the information criterion, but also to change of the values of the optimal geometric parameters of the recognition classes container. The most disturbing (disinfecting) is the sign 19, the informativeness of which is equal to $\Delta P_{19} = -0{,}15$. Removing this sign from the vocabulary also changes the optimal values of the geometric parameters of the classes of recognition containers and the exact characteristics of machine learning. But unlike in the previous case, the tendency of these indicators is the opposite.

Fig. 4 shows the graph of the averaged information criterion dependence (6) on the parameter A of the control tolerance field on the recognition signs obtained in the course of machine learning with removing the most informative sign of recognition 17 and the sign 19 that interferes.

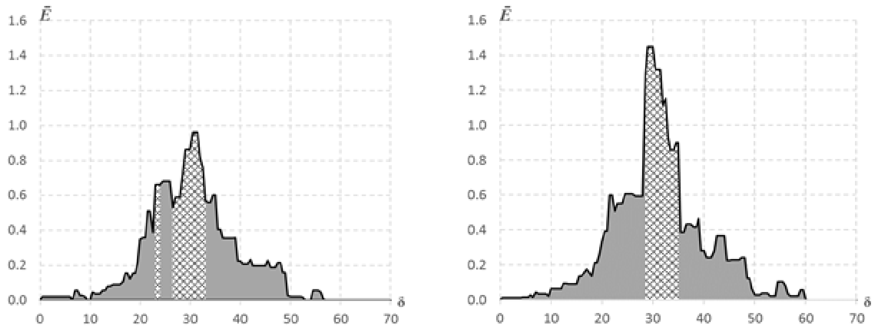

Fig. 4. Graphs of the information criterion dependence on the parameter of the field of control tolerances on the recognition signs: *a* – with an informative sign deleted; *b* – with a disturbing sign removed.

Analysis on Fig. 4 shows that comparing to the results of machine learning of the knowledge control system in the initial training matrix, there was an increase in the maximum value of the information criterion, which is equal to $\overline{E}^{*} = 1{,}45$. In addition, we received a new optimal value of the control tolerance field parameter, which was equal to $\delta^{*} = 32$. In this case, the radii of the classes of recognition containers have not changed except for the inside radius of the class container X_2^o, which became equal to $R1_2 = 13$.

6 Discussion

In this work, the difference between the maximum values of the information criteria of machine learning functional efficiency with the full set of initial tests and a set with the withdrawn test is considered as a measure of informativeness of the test. A test, the removal of which led to a decrease in the maximum value of the information criterion, was accepted as informative. As a result, 21 informative tests were selected out of 28 initial tests, for which the input learning matrix was formed according to the results of machine knowledge control. The use of informative tests set allowed to increase the maximum value of the information criterion (Fig. 4) in comparison with the use of the initial set (Fig. 2), which confirms the perspective of the proposed method.

7 Conclusions

The paper proposes an information-extreme method for evaluating the informativeness of test tasks for machine knowledge control. By eliminating from the initial set of non-informative and disinforming tests, the functional efficiency of machine learning of the knowledge control system is improved. Further improvement of the proposed method is to increase the depth of machine learning, which is determined by the number of contours of optimizing the parameters of the knowledge control system functioning in the categorical model.

8 References

1. Menqui, C., Yun L, Ying L.: Machine Learning and Modern Education. In: 4th International Conference, eLEOT 2018, Shanghai, China, pp. 41-46 (2018)
2. Lemaître, D.: Current developments in the French engineering education system. In: European Journal of Engineering Education, 42(2), pp. 131–141 (2017)
3. Zgurovskii, M. Z., Pankratova, N. D.: Sistemnyi analiz: problemy, metodologiya, prilozheniya, Kiev: Izdatelstvo Naukova dumka (2011)
4. Yu, F. T. S., Yu, E. H., Yu, A. G.: The Art of Learning: Neural Networks and Education, CRC, PressINC (2018)
5. Dovbysh, A. S., Martynenko, S. S., Kovalenko, A. S., Budnyk, M. M.: Information-extreme algorithm for recognizing current distribution maps in magnetocardiography. In: Journal of Automation and Information Sciences, 2(43), pp. 63-70 (2011)
6. Dovbysh, A. S., Budnyk, M. M., Moskalenko, V. V.: Information-Extreme Algorithm for Optimizing Parameters of Hyperellipsoidal Containers of Recognition Classes. In: Journal of Automation and Information Sciences, New York: Begell House Inc., 10 (44), pp. 35-44 (2012)