UDC 004.85.051 **doi: 10.32620/reks.2023.4.06**

Oleksandr PAPCHENKO, Borys KUZIKOV, Oksana SHOVKOPLIAS

Sumy State University, Sumy, Ukraine

ENHANCING FUNCTIONAL EFFICIENCY IN INFORMATION-EXTREME MACHINE LEARNING WITH LOGISTIC REGRESSION ENSEMBLES

*The **subject matter** of this article is the application of supervised machine learning to the task of object class recognition. The **goal** is to enhance functional efficiency in information-extreme technology (IET) for object class recognition. The **tasks** to be solved are as follows: to analyze possible ways of increasing the functional efficiency of IET approach; to implement an ensemble of models that include logistic regression for prioritizing recognition features and an IEI learning algorithm; and to compare the functional efficiency of the resulting ensemble of models on a well-known dataset with the classic approach and the results of other researchers. Methods: The method is developed within the framework of the functional approach to modeling natural intelligence applied to the problem of object classification. The following **results** were obtained: This study tries to augment existing IET to support feature prioritization as part of the object class recognition algorithm. The classical information-extreme algorithm treats all input features that are equivalently important in forming the decisive rule. As a result, the object features with strong correlation are not prioritized by the algorithm's decisive mechanism, resulting in decreased functional efficiency in the exam mode. The proposed approach solves this problem by applying a two-stage approach. In the first stage, the multiclass logistic regression applied to the input training feature vectors of the objects to be classified formed the normalized training matrix. To prevent overfitting of the logistic regression, the L2 (ridge) regularization method was used. In the second stage, the information-extreme method as input takes the result of the first stage as input. The geometrical parameters of the class containers and the control tolerances of the recognition features were considered as the optimization parameters. **Conclusions**. The proposed approach increases MNIST (Modified National Institute of Standards and Technology) dataset classification accuracy compared with the classic information-extreme method by 26.44%. The proposed approach has a 3.77% lower accuracy compared to neural-like approaches but uses fewer resources in the training phase and allows retraining of the model, as well as expanding the dictionary of recognition classes without model retraining.*

Keywords: *supervised machine learning; information-extreme machine learning; machine learning parameter regularization; algorithms ensemble; information criterion; optimization.*

1. Introduction

1.1. Motivation for research and the State of the Art

The substantial increase in data generation and accumulation has increased interest in machine learning as a whole and supervised machine learning in particular. This is generally connected to the realization that this data contains valuable hidden insights. There are three significant data science research tasks: description, prediction, and causal inference. Each of these tasks can be solved using machine learning [1]. Automatic intellectual systems, particularly Machine Learning (ML), have progressed remarkably recently. It has numerous real-life applications. For instance, ML has emerged as the method of choice for developing practical software for computer vision, speech recognition, and language processing. The particular influence of Machine Learning has also been widely felt across industries with data-intensive issues, such as consumer services, diagnosing failures in complex systems, and controlling supply chains [2].

The state-of-the-art algorithm in supervised learning includes neural-like structures – CNN (convolutional neural networks) [3]. The neural-like approach has proved to be one of the best levels of functional efficiency for object classification tasks [4]. The neural-like algorithms may contain thousands of free parameters, resulting in a high ability to learn a wide variety of different patterns. Simultaneously, this can result in high computational resource consumption during the training and exam modes. Moreover, as a result of a huge number of free parameters, these algorithms are vulnerable to overfitting; thus, regularization is usually used. There are several well-known methods [5] to decrease resource consumption and solve the problem of overfitting. One method to eliminate overfitting and fully connected neural network topology is the dropout approach, which results in a sparser network with a smaller number of parameters. This simplifies the network topology, resulting in increased speed of the exam stage and preventing overfitting. However, the method also has

ISSN 1814-4225 (print)
ISSN 2663-2012 (online)

downsides: the training stage time is increasing, and the method has several hyper-parameters that may be hard to tune, including dropout probability and learning rate.

Another popular neural-like approach includes convolutional neural networks. This method is known to be a regularized type of feed-forward neural network that learns feature engineering by applying filters or kernels [6]. Despite having one of the best classification performances, the method also has several disadvantage, including the requirement of a large number of labeled training samples for weight parameter learning [7].

The problem of high compute resource consumption is especially evident when considering the widespread interest in Internet of Things (IoT) de-vices and the technology of edge computing technology. Edge computing [8] is a distributed computing approach. The main goal of the technology is to bring compute and data servers closer to the consumer, i.e., to the location where they are needed. The main benefit of this effort is that it saves network bandwidth or reduces the latency of application request processing. For machine learning informational systems, applying this approach means moving models closer to the user.

One typical application is federated machine learning [9]. The machine learning system comprises shared models spread across decentralized nodes. The models in each node are optimized to use computational resources as efficiently as possible. The technique is also used when every node must use its training data for confidentiality purposes. The technique has been used on smartphones [10] to enhance the next-word prediction for virtual keyboards. Because the training data never leaves the node where it is executed, federated learning was previously conceived to maintain the confidentiality of the training data. In such a way, one of the prerequisites of federated learning is to have a model that enables it to run on comparatively low-powered computation nodes, such as smartphones. This study was conducted to find the model's hyper parameters, considering not only the functional efficiency of the model but also computational resource consumption. This research is based on performing multi-objective optimization solutions for finding acceptable machine learning models in terms of model accuracy and resource usage [11]. As described in the research, the proposed approach for hyperparameters optimization includes performing hyperparameter tuning directly on the target device, which is often not feasible due to the resource limitations of the target device. The hyperparameter tuning process imposes high memory requirements for evaluating multiple models on large training sets.

Thus, one of the biggest problems arising for neural-like approaches is the need for huge amounts of computational resources [12]. The scale of the problem reached a point when the carbon footprint of machine learning systems became a concern. While it is still hard to estimate the actual carbon footprint of a particular machine learning system, it is still important from an ethical point of view to decrease its impact on the environment [13]. In such a way, there are increasingly obvious trade-offs between the machine learning functional efficiency and simplicity of the model and as a result, the number of consumed resources. The promising method is information-extreme intellectual technology (IEIT). This method requires less computational power than neural-like approaches [14].

1.2. Objectives and approaches

One of the main objectives in IEIT research field is the development of approaches to improve the functional efficiency of the method.

There are several known techniques for increasing the functional efficiency of the IEIT. One of the methods is to use information-extreme machine learning on a hierarchical data structure. This approach is used to increase the functional efficiency of the IEIT, especially in the case of multiclass classification problems. The method uses the hierarchical structure of data in the form of a so-called decursive binary tree [14]. The data structure in the form of a binary tree is called decursive, in which the attribute from the top of the upper tier is transferred to its stratum top of the lower tier. The recognition class alphabet is given in the form of a decursive hierarchical structure $\{x^{o}_{h,s,m} \mid h = \overline{1,H}\,; s = \overline{1,S}; m = \overline{1,2}\}$ where H is the number of tiers; S – the number of executions on the h-th tier; "2" for m means the number of classes on each level. On each tier, the algorithm forms the binary decisive rule to select the correct object class with the highest probability. It was shown that in performing classification for the alphabet with multiple classes, the hierarchical approach has a higher probability of predicting the correct class compared with the standard IET approach [15].

Another approach to increase the functional efficiency of the IET method is feature selection preprocessing. The classic information-extreme method treats all features as equally important. In this way, the informative, non-informative, and counter-informative features exert equal influence on the results during the examination stage. The input mathematical description is sensitive to containing the most important features to produce the decisive rule with the maximum full probability of correct object class recognition [16, 17]. One of the known approaches to solve this problem is to use more features considered during class recognition, in which the number informative, non-informative, and counter-informative are amortized. The increase in the feature vector dimension may be achieved due to the introduction of synthetic features or by extending the feature space by considering more feature categories. For example, for object recognition on RGB images, one may consider not only the RGB components but also half-tone elements in the recognition of color images [18]. The obvious downside of this approach is the growing feature space. The compute resources needed for the exam mode is growing proportionally to the size of the feature space. On the other hand, approaches such as [19] attempt to address this issue by assessing the feature contribution to the overall functional efficiency of the decision-making system through a sequential reduction of the feature vocabulary. However, this significantly increases the time required for the training stage. In such a way, both approaches (with expansion and reduction feature space) might be not appropriate for systems with small computational capacity, such as IoT.

This study attempts to increase the functional efficiency of the IET approach by using logistic regression as the first stage of the algorithm. The proposed approach follows from the hypothesis that, during the process of decisive rule construction, IEIT treats all input features equally important – the most important features are not prioritized; thus, it is assumed that combining IEIT with a Machine Learning method that assigns weights to the input features in the process of learning may increase the functional efficiency compared to classical IEIT. The MNIST [20] dataset is used to verify the results of the proposed approach. The MNIST dataset was chosen because it is well known and has been applied to numerous ML approaches, which in turn enables us to compare our results with those of other algorithms for object class recognition.

2. Formalization of the proposed approach

2.1. Basic and enhanced EIT approach on the learning stage

Consider the formalized formulation of the problem of information synthesis, which can be studied within the ensemble of logistic regression and IET. The information system was developed for automatic object classification using input structured feature vectors.

Let $\{X_m^o \mid m = \overline{1,M}\}$ is alphabet recognition classes, where M is the total number of classes. Suppose a given object property training matrix $\left\|y_{m,i}^{(j)}\right\|$, $i = \overline{1,N}$, $j = \overline{1,n}$, where N, n are the number of class features recognition and implementations, respectively. It can be seen that the matrix row $\left\{y_{m,i}^{(j)} \mid i = \overline{1,N}\right\}$, determines the j-th feature vector and column $\left\{y_{m,i}^{(j)} \mid j = \overline{1,n}\right\}$ – random sampling of the i-th object feature. It's known that for the standard multiclass logistic regression object classification approach we find such weights $W = \{w_{m,i} \mid m = \overline{1,M},\ i = \overline{1,N}\}$, and biases $B = \{b_m \mid m = \overline{1,M}\}$, for which the loss function reaches the minimum. The loss function in our case is a negative log likelihood function in form: $E_{in} = \sum_{j=1}^{n} \log \hat{y}_{(W,B),j}^{(y_j)}$ – in-sample error equals the sum of the natural logarithm of predicted probability for true label y_j over all test samples, where j is the sample index [21]. Formally we have the minimization problem: $\underset{W,B}{\operatorname{argmin}} E_{in}$. The logistic regression is parametrized with three hyperparameters: n_{it} – number of iterations, lr – learning rate, λ – L2 regularization parameter for penalty term of the loss function, values for these parameters are determined experimentally [22]. The logistic regression takes as input the structured features vector $\{y_i \mid i = \overline{1,N}\}$ and transforms it to the probability distribution vector $\{p_m \mid m = \overline{1,M}\}$, notice the probability distribution vector dimension equal to the number of recognition classes. In proposed approach the input object property training matrix $\left\|y_{m,i}^{(j)}\right\|$, $i = \overline{1,N}$, $j = \overline{1,n}$ is transformed to the new training matrix $\left\|p_{m,i}^{(j)}\right\|$ the logistic regression parametrized by optimal weights and biases found in the training stage for each row of the initial matrix. In this way obtained, a new input property training matrix $\left\|p_{m,i}^{(j)}\right\|$, $i = \overline{1,M}$, $j = \overline{1,n}$. The IEI technology approach transforms the input training matrix Y into a training binary matrix X, which adapts to the maximum possible probability of making correct classification decisions by the method of permissible transformations in machine learning. For Hamming binary space introduced set of $\{g_m\}$ machine learning parameters that affect the functional efficiency of the IET algorithm. The set of optimal parameters is represented in the form:

$$g_m = < x_m, d_m, \delta_k >, \qquad (1)$$

where x_m – is the average features structured vector of the recognition class from the alphabet X_m^o; d_m – radius of the recognition class hyper spherical container for X_m^o, which is restored in the radial basis of the recognition features space; δ_k – parameter of the control tolerances for the recognition features field, which is equal to half of the symmetric field control tolerances.

Required:

– to determine machine learning parameters (1), which provide the maximum of the averaged information criterion:

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

ISSN 1814-4225 (print)
ISSN 2663-2012 (online)

where $E_m^{(k)}$ is the functional efficiency information criterion in recognition of the implementation of class X_m^o, calculated on the k-th step of the algorithm; G_E – admissible working area for optimization of the information functional efficiency criterion, $\{k\}$ – machine learning algorithm learning steps;

– form a training matrix, by transforming the prior classified fuzzy partition $\tilde{\mathcal{R}}^{|M|}$ to subperceptual binary Hamming space;

– evaluate the Machine Learning functional efficiency to decide whether the implementation of the recognized image belongs to one of the classes of the alphabet.

In this way, the learning process consists of optimizing parameters (1) according to the information criterion (2).

Consider the categorical functional model of Machine Learning. The model is presented in the form of a directional graph of mapping by operators of the corresponding sets used in the learning process. Such representation enables better visualization of the learning procedure. The input mathematical description is presented in the following structure:

$$I =< G, T, Z, \Omega, Y, P, X; f_1, f_2, f_3, f_4 >$$

where G is the space of influence factors that affect the functioning of the automatic recognition system; T − set input data (training dataset); Z – space of possible recognition classes; Ω – space of recognition features; Y – input training matrix; P – normalized training matrix;

X – training binary training matrix; f_1 – the operator of recognition features analysis; f_2 – operator formation of the training matrix Y; f_3 – operator to convert the input training matrix to normalized training matrix P by applying the logistic regression; f_4 – operator for converting the input training matrix Y to the training matrix X defined in the Hamming space.

Notice, the first stage of the proposed ensemble is presented in this model as operator f_3. Fig. 1 shows a categorical functional model in the form of a directed graph of an ensemble of logistic regression and an information-extreme machine learning system for object class recognition with optimization of control tolerances for recognition features.

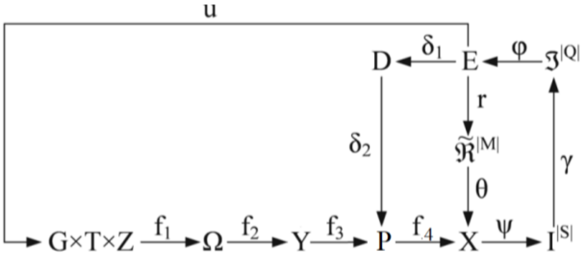

Fig. 1. Categorical functional model of machine learning

In Fig. 1 the term E consists of information criterion value (2) calculated at each step. This is common to all optimization contours for the parameters of vector (1). Operator $r: E \rightarrow \tilde{\mathcal{R}}^{|M|}$ in the machine learning process restores the radial basis of the binary feature space containers of recognition classes. In turn operator θ maps the partition on the fuzzy distribution of a priori classified binary vectors of recognition class features $\tilde{\mathcal{R}}^{|M|}$. In the following step of the contour the operator ψ performs the transition $X \rightarrow I^{|S|}$, where $I^{|S|}$ is the set of hypotheses $\gamma_1: x_m^{(j)} \in X_m^o$.Operator γ determines the set of accuracy characteristics $\mathfrak{I}^{|Q|}$ and operator φ calculates the set of values E of the information optimization criterion, which is functional from the accuracy characteristics. Next, the categorical model contains the contour for optimization of the control tolerance parameter closed by a term set D of allowable values for the system of control tolerance. The operator δ_1 at each step of machine learning changes the control field, and the operator δ_2 evaluates the dependence of the recognition features of a given control field on the tolerances. Operator u regulates the machine learning process.

The main difference between the proposed model and the classical IET approach is the existence of operation f_3 which transforms the input object-feature training matrix X into normalized training matrix P obtained by the application of logistic regression to the original training matrix realizations.

In our case, the information-extreme machine learning algorithm with optimization of the control tolerance parameter for features of recognition corresponds to the second level of machine learning depth. It is presented as a two-cycle iterative procedure for determining the global maximum information optimization criterion (2). The internal cycle implements the basic approach of information-extreme machine learning. In this cycle, we determined the global maximum of information criterion (2), thus finding the optimal radii of recognition classes of hyper spherical containers. In our case the implementation of control tolerances is applied in parallel manner, meaning that all tolerances for recognition features change simultaneously by a given value. The input to the machine learning algorithm of the information-extreme stage is normalized matrix $\left\|p_{m,i}^{(j)}\right\|$ obtained by applying logistic regression.

1. The main stages of information-extreme machine learning: Determine the optimal values of weights W and biases B of the logistic regression model by running the logistic regression training procedure on the input training matrix $\left\|y_{m,i}^{(j)}\right\|$, $i = \overline{1,N}$, $j = \overline{1,n}$, Apply the first stage of the machine learning ensemble by applying the logistic regression model parameterized by optimal weights W and biases B to the input training

matrix. As a result, obtain the normalized training matrix $\left\|p_{m,i}^{(j)}\right\|$. Each row of the normalized matrix is the probability distribution of the recognition classes obtained from application of softmax function.

2. Calculating the training matrix for the primary recognition class, denoted as X_1^o, involves determining the average feature vector $\{p_{1,i} | i = \overline{1,N}\}$ for control tolerance. Notice that in the previous step, logistic regression was applied, and the number of features in the normalized matrix is equal to the number of recognition classes.

3. Determine the binary vectors of recognition class X_1^o, using the following rule:

$$x_{1,i}^{(j)} = \begin{cases} 1, \text{ if } p_{1,i} - \delta_i \le p_{1,i}^{(j)} \le p_{1,i} + \delta_i, \\ 0, \text{ if else;} \end{cases}$$

4. Form average binary vectors for each class implementations by the following rule:

$$x_{m,i} = \begin{cases} 1, \text{if } \frac{1}{n}\sum_{j=1}^{n} x_{m,i}^{(j)} > \rho_m, \\ 0, \text{if else,} \end{cases}$$

where ρ_m is a selection level hyperparameter. Possible values are in the range (0, 1].

5. Set division of average feature vectors according to the rule of "nearest neighbours" $\mathcal{R}_m^{|2|} = < x_{m,} x_1 >$, where x_1a binary average vector of neighbouring class is carried out according to the following scheme:

a) form the structuring vectors set $\{x_m\}$ starting from the basis class vector x_1;

b) the matrix constructed with dimensions M×M. In each cell of the matrix is the code distance between the corresponding two recognition classes;

c) find the minimum for each row of the matrix formed in the previous step;

d) form a structured set of pairwise partitioning $\{\mathcal{R}_m^{|2|} | m = \overline{1,M}\}$;

e) find code distances d_m for each recognition class, for which the information criterion has the maximum value. The restriction for radii d_m of spherical container is: $d_1 < d(x_1 \oplus x_2) - 1$. In other words, the distance between neighbours should be greater than the container radius;

f) implement the procedure (2), and find the optimal control tolerance.

6. STOP.

After the procedure is executed, we have optimal weights W and biases B for the logistic regression model and IET optimal parameters $\{x_m^*\}$– etalon vectors for each class, $\{d_m^*\}$– the recognition classes containers radii, and system of control tolerances $\{\delta_i^*\}$ on recognition features.

As a criterion for optimization of machine learning parameters IET is known to use the modified Kullback measure.

$$E_m^{(k)} = \frac{n - \left(K_{1,m}^{(k)} + K_{2,m}^{(k)}\right)}{n} \log_2 \frac{2n + \xi - K_{1,m}^{(k)} - K_{1,m}^{(k)}}{K_{1,m}^{(k)} + K_{1,m}^{(k)} + \xi}$$

where $K_{1,m}^{(k)}$ – number of false negative events, $K_{2,m}^{(k)}$ – number of false positive events, ξ – small number to prevent division on zero.

Having optimal parameters obtained in the learning stage as the next stage built decisive rules in form:

$$\left(\forall X_m^o \in \tilde{\mathcal{R}}^{|M|}\right)\left(\text{if } \left((\mu_m > 0) \,\&\, \mu_m = \max_{\{m\}}\{\mu_m\}\right), \text{ then } x^{(j)} \in X_m^o, \text{ else } x^{(j)} \notin X_m^o\right), \quad (3)$$

where μ_m is a membership function:
$\mu_m = 1 - d(x_1 \oplus x_2)/d_m^{\cdot}$, where $d(x_m^* \oplus x^j)$ is a code distance between the etalon and classified vectors.

2.2. System operation in the exam mode

After forming the decisive rules, the system is ready to function in the exam mode. The functional efficiency of the system is possible to measure in the exam mode. The categorical functional model is presented in the form of a directed graph and has one contour. In the categorical model (Fig. 2) operator u_E regulates the process of system functioning in the exam mode; Sets G, Z, Ω and f_1 have the same sense as previously, and T – contains input data for recognition.

$$\begin{tikzcd} G\times T\times Z \arrow[r, "f_1"] & \Omega \arrow[r, "f_2"] & y \arrow[r, "f_3"] & p \arrow[r, "f_4"] & x \arrow[r, "f_5"] & \mathfrak{R}^* \arrow[r, "\psi_1"] & F \arrow[r, "\psi_2"] & I^{|M+1|} \arrow[lllllll, bend right, "u_E"'] \end{tikzcd}$$

Fig. 2. Categorical model of system functioning in the exam mode

Let us clarify the meaning of some operators according to the change in set T. Operator f_2 forms an examination vector of recognition features, similar in structure to the vectors of the training matrix. Similar to the training mode, operation f_3 applies logistic regression to the input feature vector, forming a probability distribution vector size equal to the number of recognition classes. Operator f_4 generates a binary vector x according to the optimal control tolerances obtained during the machine learning stage. Operator f_5 maps this vector for the optimal division of $\mathfrak{R}^*$ recognition classes.

ISSN 1814-4225 (print)
ISSN 2663-2012 (online)

Operator ψ_1 calculates the value of the decisive rule (3) and forms a term set F, and operator ψ_2 on the maximum value of the decisive rule determines the affiliation of the vector x to one of the alphabets classes (verification hypotheses $I^{|M+1|}$). The set of possible hypotheses $I^{|M+1|}$ contains additional hypothesis γ_{M+1} which is accepted in the case of system failure to classify input data.

3. Experimental result and discussion

The above two-stage algorithm was implemented to recognize handwritten digits from the MNIST dataset. The MNIST database contains 60000 training images of handwritten digits and 10000 testing images. The alphabet of recognition classes has size of 10 – one for each digit. Each image has 28×28 grayscale resolution. No preprocessing was applied to the images. The Fig. 3 shows example images for each recognition class from the dataset.

Fig. 3. Example images from the MNIST dataset

According to the above algorithms in the first stage, the logistic regression model was trained. The following hyperparameters were used: number of iterations n_{it} – 2000, learning rate lr – 0.05, λ – regularization parameter 0.1. As a result of the logistic regression training stage, the set of optimal weights W and biases B were found. The Fig. 4 shows the visualization of optimal weights for each digit class.

Next, to the training data, logistic regression with optimal parameters was applied. As a result, we obtained a normalized training matrix consisting of probability distribution vectors over the alphabet of recognition classes. Since, the values of the normalized training matrix represent probabilities – its values are in the range [0, 1], for convenience the values were scaled to have values in range [0, 256]. The normalized training matrix was used for the second stage, i.e., the determination of optimal IET parameters. The working area for the control tolerances machine learning parameter δ was took roughly the half of the range of possible features values. Because in our case features are in range [0, 256] – the working area was taken 120.

From Fig. 5 it is visible that for the classic IET algorithm maximum information criterion $E = 0.24$ is reached at the value of control tolerance $\delta = 5$.

From Fig. 6 it is visible that for the ensemble of logistic regression and IET algorithms, the maximum information criterion $E = 0.91$ is reached at the value of control tolerance $\delta = 90$. The value of the information criterion for the ensemble approach is noticeably higher compared with the classic IET approach. On graphs, the area filled with gray color represents the control working area. The area filled with dark gray represents the area where the information criterion reaches maximum.

To construct the decisive rules (3), the optimal control tolerance parameter δ^{*} was found optimal values for hyper spherical containers of recognition classes. As shown in table 1 and 2, all classes in the enhanced approach become more compact, which led to a decrease in type II error at the examination stage (see fig 7,8).

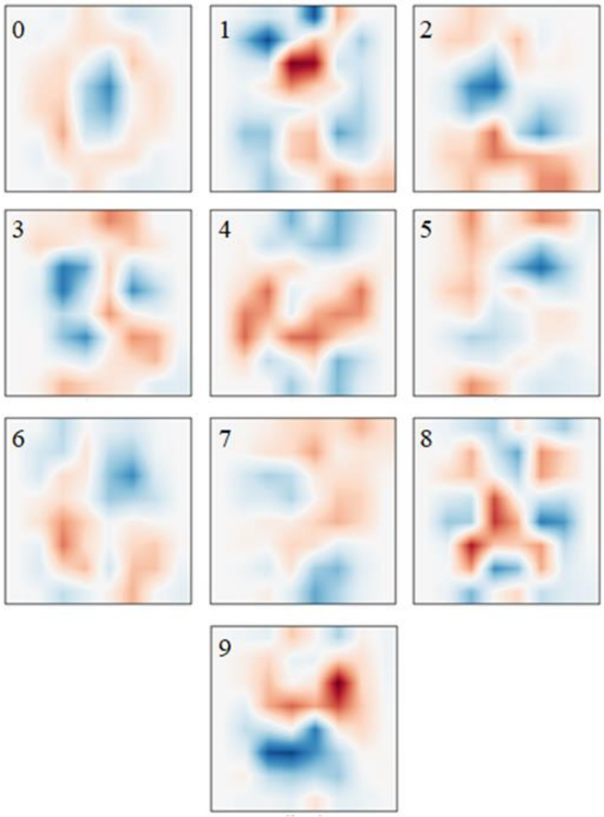

Fig. 4. Visualization of logistic regression optimal weights, 0-9 are label for corresponding classes

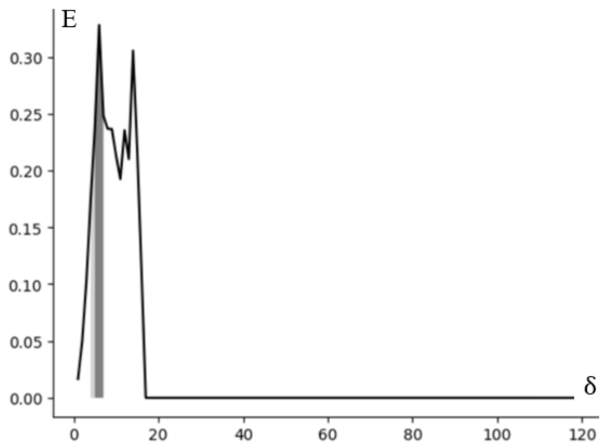

Fig. 5. Dependence graph of the information optimization criterion (E) from the control tolerance (δ) field parameter for IET approach

In the examination mode, we evaluated the functional efficiency of the method. We used the test sub dataset from MNIST dataset. For IET approach the accuracy is 69.56%. For the proposed ensemble of logistic and IET, the accuracy is 96.0%. We can see that the accuracy increase is 26.44%. For reference, the state-of-the-art neural-based approach has an accuracy of 99.77% [23, 24]. The proposed approach has 3.77% lower accuracy; however, it is characterized by lower requirements for computational resources.

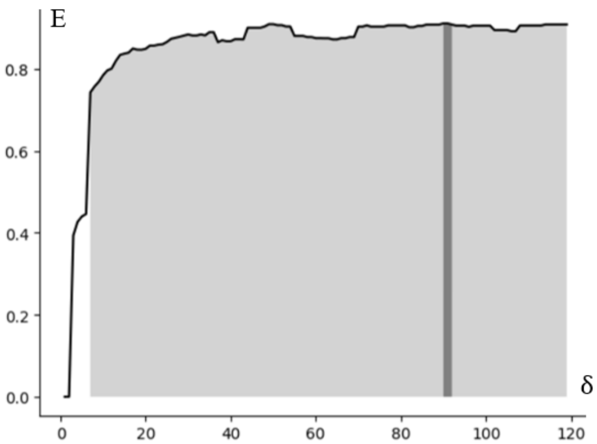

Fig. 6. Dependence graph of the information optimization criterion from the control tolerance (δ) field parameter for the ensemble of logistic regression and IET

Table 1

Optimal radii for hyper spherical containers using classic IET and enhanced approach

Classes	0	1	2	3	4	5	6	7	8	9
Radii (classic IET)	8	10	13	10	13	11	10	11	12	8
Radii (enhanced approach)	1	2	1	2	1	2	2	2	2	2

The Fig. 7 shows the probability distribution of the predicted class using IET approach. This shows that the model has severe misclassification for some classes. For example, for class with digit 1, the model correctly classified only 44% of the input samples.

Actual label \ Predicted label	0	1	2	3	4	5	6	7	8	9
0	1	0	0	0	0	0	0	0	0	0
1	0	0.44	0.23	0.02	0.07	0	0.02	0	0.21	0
2	0	0	0.77	0.02	0	0.02	0	0.07	0.11	0
3	0	0.02	0.09	0.76	0	0.02	0	0	0.11	0
4	0	0	0	0	0.84	0	0	0.11	0.05	0
5	0.02	0.02	0.1	0.06	0.02	0.71	0.02	0	0.02	0.02
6	0	0	0	0	0.15	0	0.81	0	0.04	0
7	0	0.04	0.04	0	0.06	0.02	0	0.81	0.02	0
8	0	0.08	0.04	0	0.04	0.08	0	0.02	0.73	0
9	0.02	0.02	0.11	0.06	0.04	0.11	0.02	0.02	0.15	0.45

Fig. 7. Probability distribution of the predicted classes for IET approach

The Fig. 8 shows the probability distribution of the predicted classes using an ensemble of logistic regression and IET. It shows that the class for digit 1 still has the lowest correct classification rate – 86 %, however it is obvious that the false positive errors for each class are generally spread across two to three classes – which in turn gives hope that hierarchical IET may have a positive effect on increasing the functional efficiency of the ensemble approach.

Actual label \ Predicted label	0	1	2	3	4	5	6	7	8	9
0	1	0	0	0	0	0	0	0	0	0
1	0	0.86	0	0	0	0	0.02	0	0.04	0.07
2	0	0	0.98	0	0	0	0	0	0.02	0
3	0	0	0	0.96	0	0	0	0	0.02	0.02
4	0	0	0	0	0.97	0	0	0.03	0	0
5	0	0	0	0	0	0.98	0	0	0	0.02
6	0	0.02	0	0	0	0	0.98	0	0	0
7	0	0	0	0	0	0	0	1	0	0
8	0	0.04	0	0	0	0.02	0	0	0.94	0
9	0	0	0	0	0	0.04	0	0	0.02	0.94

Fig. 8. Probability distribution of the predicted classes for the ensemble of logistic regression and IET

As noted before, the result of the softmax function from the first stage of the ensemble was scaled to have values in the range [0, 256]. This was done to provide a convenient way to compare results from the classical IET approach and ensemble approach. We studied the influence of input feature absolute values scalers on functional efficiency and performance of the training stage. On decreasing the scaled feature area to range [0, 120] and keeping the learning step of control tolerance equal to 1, the functional efficiency did not decrease, and the training stage took 48.5% less time compared to the feature value range [0, 256]. However, on further decreasing the feature value area and keeping the same control tolerance range, significantly decreasing the functional efficiency and converging to the accuracy of plain logistic regression 95.11 %. For instance, by decreasing the feature value range to [0, 100], the accuracy drops to 95.33 %. This can be explained by the fact that by decreasing the range of feature values and maintaining the same learning rate, the control tolerance resolution of the traversed search area is also decreased.

Conclusions

Research was performed to increase the functional efficiency of information-extreme machine learning methodology applied to the task of object class recognition. Formed hypothesis: Because the information extreme approach treats all input features equally important, the ensemble of multi-class logistic regression and IET increases functional efficiency compared to

ISSN 1814-4225 (print)
ISSN 2663-2012 (online)

classical IET. The initial hypothesis was verified using the MNIST dataset. The testing results show an increase in the recognition accuracy of the proposed method compared with the classical information extreme approach of 26.44 %.

The proposed approach for object classification has an acceptable rate of functional efficiency while having much lower computational costs than neural-like approaches. The alternative approach to increasing the functional efficiency of IET is to use a hierarchical approach instead of a linear one [15]. This paves the way for further improvements in the functional efficiency of machine learning by optimizing additional system parameters and using ensemble logistic regression and hierarchical IET.

In summary, the proposed ensemble has the following advantages:

– compared with the classic IET approach, increased functional efficiency of object classes recognition on 26.44% of MNIST dataset;

– flexibility to retrain them through expansion of the recognition classes alphabet;

– compared with the neural-like approach, it requires much less computational data.

Future research will continue to solve the problem of increasing the functional efficiency of IET under limited computing resources. The directions of the automated reduction of the dictionary of features and the reduction of the variability of the values of the recognition features seem promising. The power of these parameters is a multiplier in the number of cycles for optimizing the geometric parameters of the system of control tolerances of recognition features, so they directly affect the running time of the learning algorithm. The complementary research direction is adopting an ensemble of logarithmic regression and IET models with a hierarchical approach to building a classifier – it is a perspective approach to enhancing the results of the exam stage.

Contributions of authors: conceptualization, methodology – **Oleksandr Papchenko**, **Borys Kuzikov**; program realization of algorithm, original draft preparation – **Oleksandr Papchenko**; verification, review, and editing – **Borys Kuzikov, Oksana Shovkoplias**.

All the authors have read and agreed to the published version of this manuscript.

References

1. Jiang, T., Gradus, J. L., & Rosellini, A. J. Supervised machine learning: a brief primer. *Behavior therapy*, 2020, vol. 51, iss. 5, pp. 675-687. DOI: 10.1016/j.beth.2020.05.002.
2. Pugliese, R., Regondi, S., & Marini, R. Machine learning-based approach: global trends, research directions, and regulatory standpoints. *Data science and management*, 2021, vol. 4, pp. 19-29. DOI: 10.1016/j.dsm.2021.12.002.
3. Sarker, I. H. Machine learning: algorithms, real-world applications and research directions. *SN computer science*, 2021, vol. 2, no. 3, article no 160, 21 p. DOI: 10.1007/s42979-021-00592-x.
4. Ciresan D., Meier, U., & Schmidhuber J. Multi-column deep neural networks for image classification. *2012 IEEE conference on computer vision and pattern recognition (CVPR)*, Providence, RI, 2012, pp. 3642-3649. DOI: 10.1109/cvpr.2012.6248110.
5. Srivastava, N., Hinton, G., Krizhevsky, A., Sutskever, I., & Salakhutdinov, R. Dropout: a simple way to prevent neural networks from overfitting. *Journal of machine learning research*, 2014, vol. 15, no. 56, pp. 1929-1958. Available at: http://jmlr.org/papers/v15/srivastava14a.html (accessed: 28.08.2023).
6. Albawi, S., Mohammed, T. A., & Al-Zawi, S. Understanding of a convolutional neural network. *2017 international conference on engineering and technology (ICET)*, Antalya, Turkey, 2017, pp. 1-6. DOI: 10.1109/ICEngTechnol.2017.8308186.
7. Han, D., Liu, Q., & Fan, W. A new image classification method using CNN transfer learning and web data augmentation. *Expert systems with applications*, 2018, vol. 95, pp. 43-56. DOI: 10.1016/j.eswa.2017.11.028.
8. Shi, W., Cao, J., Zhang, Q., Li, Y., & Xu, L. Edge computing: vision and challenges. *IEEE internet of things*, 2016, vol. 3, iss. 5, pp. 637-646. DOI: 10.1109/jiot.2016.2579198.
9. Yang, Q., Liu, Y., Chen, T., & Tong, Y. Federated machine learning. *ACM transactions on intelligent systems and technology*, 2019, vol. 10, iss. 2, article no. 12. 19 p. DOI: 10.1145/3298981.
10. Hard, A., Rao, K., Mathews, R., Ramaswamy, S., Beaufays, F., Augenstein, S., Eichner, H., Kiddon, C., & Ramage, D. *Federated learning for mobile keyboard prediction*. Available at: https://arxiv.org/abs/1811.03604 (accessed: 28.08.2023). DOI: 10.48550/arXiv.1811.03604
11. Preuveneers, D., Tsingenopoulos, I., & Joosen, W. Resource usage and performance trade-offs for machine learning models in smart environments. *Sensors*, 2020, vol. 20, iss. 4, article no. 1176. 27 p. DOI: 10.3390/s20041176.
12. Bouguettaya, A., Kechıda, A., & Taberkıt, A. M. A Survey on Lightweight CNN-Based Object Detection Algorithms for Platforms with Limited Computational Resources. *International Journal of Informatics and Applied Mathematics*, 2019, vol. 2, iss. 2, pp 28-44. Available at: https://dergipark.org.tr/en/pub/ijiam/issue/52418/654318. (accessed: 28.08.2023).

13. Tamburrini, G. The AI carbon footprint and responsibilities of AI scientists. *Philosophies,* 2022, vol. 7, iss. 1, article no. 4. DOI: 10.3390/ philosophies7010004.

14. Dovbysh, A. S., Shelehov, I. V., Romaniuk, A. M., Moskalenko, R. A., & Savchenko, T. R. Decision-Making support system for diagnosis of breast oncopathologies by histological images. *Cybernetics and systems analysis*, 2023, vol. 59, iss. 3, pp. 493-502. DOI: 10.1007/s10559-023-00584-0.

15. Dovbysh, A. S., & Piatachenko, V. Y. Hierarchical clustering approach for information-extreme machine learning of hand brush prosthesis. *Proceedings of the 5th international conference on computational linguistics and intelligent systems,* Kharkiv, 2021, pp. 1706-1715.

16. Naumenko, I., Myronenko, M., & Savchenko, T. Information-extreme machine training of on-board recognition system with optimization of RGB-component digital images. *Radioelectronic and computer systems*, 2021, no. 4, pp. 59-70. DOI: 10.32620/reks.2021.4.05.

17. Dovbysh, A., Liubchak, V., Shelehov, I., Simonovskiy, J., & Tenytska, A. Information-extreme machine learning of a cyber attack detection system. *Radioelectronic and computer systems*, 2022, no. 3, pp. 121-131. DOI: 10.32620/reks.2022.3.09.

18. Dovbysh, A., Shelehov, I., Prylepa, D., & Golub, I. Information synthesis of adaptive system for visual diagnostics of emotional and mental state of a person. *Eastern-European Journal of Enterprise Technologies*, 2016, vol. 4, iss. 9(82), pp. 11-17. DOI: 10.15587/1729-4061.2016.75683.

19. Dovbysh, A., Shelehov, I., Pylypenko. S., & Berest, O. Estimation of Informativeness of Recognition Signs at Extreme Information Machine Learning of Knowledge Control System. *CEUR Workshop Proceedings*, 2019, article no. 2362. 10 p.

20. Deng, L. The MNIST database of handwritten digit images for machine learning research. *IEEE signal processing magazine*, 2012, vol. 29, iss. 6, pp. 141–142. DOI: 10.1109/msp.2012.2211477.

21. Mehta, P., Bukov, M., Wang, C.-H., Day, A. G. R., Richardson, C., Fisher, C. K., & Schwab, D. J. A high-bias, low-variance introduction to Machine Learning for physicists. *Physics reports*, 2019, vol. 810, pp. 1-124. DOI: 10.1016/j.physrep.2019.03.001.

22. Ng, A. Y. Feature selection, L1 vs. L2 regularization, and rotational invariance. *Proceedings of the twenty-first international conference on Machine learning (ICML '04)*, New York, NY, USA. 78 p. DOI: 10.1145/1015330.1015435.

23. Baldominos, A., Saez, Y., & Isasi, P. A Survey of Handwritten Character Recognition with MNIST and EMNIST. *Applied Sciences*, 2019, vol. 9, iss. 15, article no. 3169. DOI: 10.3390/app9153169.

24. Islam, M. A., & Iacob, I. E. Manuscripts Character Recognition Using Machine Learning and Deep Learning. *Modelling,* 2023, vol. 4, iss. 2, pp. 168-188. DOI: 10.3390/modelling4020010.

Received 17.09.2023, Accepted 20.11.2023

ПІДВИЩЕННЯ ФУНКЦІОНАЛЬНОЇ ЕФЕКТИВНОСТІ ІНФОРМАЦІЙНО-ЕКСТРЕМАЛЬНОГО МАШИННОГО НАВЧАННЯ ЗА ДОПОМОГОЮ АНСАМБЛІВ ЛОГІСТИЧНОЇ РЕГРЕСІЇ

Олександр Папченко, Борис Кузіков,
Оксана Шовкопляс

Об'єктом дослідження є застосування керованого машинного навчання для задачі класифікації об'єктів. **Метою** дослідження є підвищення функціональної ефективності інформаційно-екстремальної технології (IET) машинного навчання. **Задачі** дослідження: проаналізувати можливі шляхи підвищення функціональної ефективності навчання моделей за IET; реалізувати ансамбль моделей що включають логістичну регресію для *пріоритезації* ознак розпізнавання та алгоритм навчання за IEI; порівняти функціональну ефективність запропонованого підходу із класичним та результатами інших дослідників на відомому наборі даних. **Методи**. Метод розроблено в рамках функціонального підходу до моделювання природного інтелекту в застосуванні до задачі класифікації об'єктів. Отримані такі **результати**. Запропоновано метод покращення IET шляхом додавання ваги ознак розпізнавання. Класичний алгоритм розглядає всі вхідні ознаки як рівноправні при формуванні вирішального правила. У результаті інформативні ознаки розпізнавання не пріоритезуються алгоритмом побудови вирішальних правил у порівнянні із не-інформативними чи контрінформативними, що призводить до зниження функціональної ефективності в режимі екзамену. Запропоновано двоетапний підхід, де на першому етапі до навчальної вибірки застосовується багатокласова логістична регресія – формується нормалізована навчальна матриця. Для запобігання перенавчання логістичної регресії використовувався метод регуляризації L2 (RIDGE). На другому етапі застосовується інформаційно-екстремальний метод навчання. Геометричні параметри контейнерів класів розпізнавання та контрольні до-

ISSN 1814-4225 (print)
ISSN 2663-2012 (online)

пуски на ознаки розпізнавання використовуються як параметри оптимізації етапу навчання моделі. Тестування отриманих результатів проведено на наборі даних MNIST (Modified National Institute of Standards and Technology), що дозволяє порівняти отриманий результат із результатами інших дослідників. **Висновки**. Запропонований метод підвищує точність класифікації на наборі даних MNIST на 26.44% порівняно з класичним інформаційно-екстремальним методом. Запропонований підхід має на 3.77% меншу точність у порівнянні із нейроподібними підходами, але використовує менше ресурсів на етапі навчання та дозволяє проводити донавчання моделі, а також проводити розширення словника класів розпізнавання без повного перенавчання.

Ключові слова: навчання з учителем; інформаційно-екстремальне машинне навчання; параметри машинного навчання; ансамбль алгоритмів, інформаційний критерій; оптимізація.

Папченко Олександр Іванович– здобувач каф. комп'ютерних наук, Сумський державний університет, Суми, Україна.

Кузіков Борис Олегович – канд. техн. наук, старш. викл. каф. комп'ютерних наук, Сумський державний університет, Суми, Україна.

Шовкопляс Оксана Анатоліївна – канд. фіз.-мат. наук, старш. викл. каф. комп'ютерних наук, Сумський державний університет, Суми, Україна.

Oleksandr Papchenko – PhD Student, Computer Science Department, Sumy State University, Sumy, Ukraine,
e-mail: olpapchenko@gmail.com, ORCID: 0009-0005-9918-566X, ResearcherID: JMC-1046-2023.

Borys Kuzikov – PhD, Senior Lecturer at the Computer Science Department, Sumy State University, Sumy, Ukraine,
e-mail: b.kuzikov@dl.sumdu.edu.ua, ORCID: 0000-0002-9511-5665, Scopus Author ID: 55653809800.

Oksana Shovkoplias – PhD, Senior Lecturer at the Computer Science Department, Sumy State University, Sumy, Ukraine,
e-mail: o.shovkoplyas@mss.sumdu.edu.ua, ORCID: 0000-0002-4596-2524, Scopus Author ID: 55647364100.