M. V. Dratovanyi; V. B. Mokin, Dr. Sc. (Eng.), Professor

INTELLIGENT METHOD WITH THE REINFORCEMENT OF THE SYNTHESIS OF OPTIMAL PIPELINE OF THE DATA PRE-PROCESSING OPERATIONS IN THE MACHINE LEARNING PROBLEMS

The paper is devoted to the synthesis and optimization of the pipelines of the data pre-processing operations in the problems of the machine learning models construction. It is noted that it is important to optimize the triad of these pipelines - select optimal sequence of the optimal operations with the optimal parameters. In this case, the change of even one element immediately influences the choice of all other elements and their parameters. In general case, there exists a great number of the admissible variants of such pipelines for each model of machine learning and input data (random values or time series) and, as a rule, there is no marked datasets of model training for the synthesis of such pipelines. The survey of the known approaches to the solution of such problems has been carried out, the conclusion that the best way is to formalize them as the problems of reinforcement machine learning has been substantiated t. Typical approaches to the formalization and intellectual methods of similar problems solution have been presented.

It is noted that the solution of the problems with reinforcement, as a rule, is complicated due to large dimensionality of the possible sets of the types and subtypes of the operations with different parameters and has problems with the coincidence to really optimal value during limited time. That is why, several improvements, enabling to solve this problem at certain conditions, are suggested. First, it is suggested to allocate variable and constant sections of the pipeline of the data pre-processing operations. It is also suggested for different types of the machine learning models what operations should be referred to the first and last unchangeable links and what operations – to variable link and only to this link it is suggested to apply reinforcement learning. Secondly, the algorithm of the initial setting of RL-policy parameters depending on certain statistical and other characteristics of the input data is suggested. The proposed improvement of the method with the reinforcement of the synthesis of the optimal pipeline of the operations can be applied not only for pre-processing operations but for other problems with the similar data formalization and problem set up.

Key words: *synthesis of the optimal pipeline of the operations, data pre-processing, game algorithms, ment machine learning, intelligent technology.*

Introduction

Models of the machine learning became widely used in all the spheres of human activity but in numerous problems the accuracy of the solution remains non sufficiently high. As a rule, these are problems, dealing with the processes prediction, which are non-stationary time series, for instance, those, which are influenced by the meteorological factors: concentration of the allergenic plants pollen or mushrooms spores, spreading of the virus diseases among the population, change of water quality in rivers or quality of atmospheric air, moisture content in the cereal crops, etc. [1 – 3].

Numerous papers are devoted to the study of the ways of improving the models of machine learning and their sets but less attention is paid to the stage of data pre-processing. However, without the efficient filtration of erroneous and abnormal data it is difficult to achieve high accuracy of the prediction of the basic statistical set of data. For instance, the cost of the used motor vehicle of 2,5 bn or 0 USD [4] could hardly allow to construct the efficient model for the prediction of this cost in the USA, if these data were not filtered. Greater part of the machine learning model require the independence of the features and their distribution according to the normal law, but in practice this happens seldom, especially – the second. Sometimes, data logarithmation helps, when smaller values are more frequent. One of the simplest ways of providing the stationarity of the time series (random value, changing in time) is to pass from the modeling of its values to the of modeling their increments (differences of the first order) or – to increment of these increments(differences of the second order). Also there are more complicated methods. As a rule, the sequence of the operations is called “pipeline” [5].

Unfortunately, there does not exist single agreed by all complex of the rules of the operations sequence, to maximize the accuracy of the model, which will process them. Transition to the first difference introduces changes almost in all regularities, which are to be analyzed anew. Logarithmation of the data or other nonlinear transformation requires reconsideration of the techniques to provide the conditions of stationarity, as all the differences can be subjected to changes. Besides, greater part of the pre-processing operations have the parameters: for the filtration of the anomalies it is sufficient to know the upper threshold (as a rule, certain quantile – P95, P90 or quantile Q3=P75 or other) and low threshold (P05, P10 or Q1=P25 or other), for taking the differences - their order, for nonlinear transformations of data, if it is not a simple logarithmation, also it is necessary to know certain parameters. That is, it is necessary to identify optimal triad for the set pipeline to find optimal sequence of optimal operations with optimal parameters, which optimize certain criterion of optimality (matrix) of the machine model which further will process data at the output of this pipeline.

There exists a number of approaches to the construction of such pipelines. As a rule, expert intuitive approaches or optimization on the base of the oriented acyclic graph are used or various neural network design. In general, learning methods with a teacher [6], multiagent optimization [7] or simple Q-learning with the reinforcement are used [8]. However, all these methods have such drawbacks: 1) small number of admissible methods of preprocessing, that often leads to low efficiency of the preprocessing on the whole; 2) large dimensionality of the adjacency matrices for the selection of optimal strategy of choosing the methods of the preprocessing of the data, that leads to the considerable duration of the program operation; 3) these methods often require the operators or experts interference, thus – the universality of the technology is not provided and there exists the risk of obtaining only quasi optional solutions regarding the improvement of the data prediction accuracy (for time series – forecasting accuracy).

The **aim** of the given research is improvement of the methods of the automatic identification of the optimal pipeline operations of the data preprocessing by means of improvement the reinforcement learning methods for the enhancement of accuracy and rate of prediction, using the models of the machine learning.

Survey of the preprocessing operations of the machine learning models

Various researchers distinguish numerous preprocessing operations of the machine learning models. As a rule, they are differentiated by the types [6 – 9] (we will denote as the elements of the set A of all the possible actions):

- Clearing (A_0):
  - Imputing (A_{00}):
    - average between neighbours (A_{000});
    - methods of the matrices factorization (MF) (A_{001});
    - multifactoral imputing by means of chain-equations (MICE) (A_{002});
    - method of K-Nearest Neighbours (KNN)) (A_{003}) etc.;
  - Deduplication (A_{01}):
    - exact detection and deduplication (ED) (A_{010});
    - approximate detection and deduplication (AD) (A_{011}) etc.;
  - inconsistency and erroneous data detection (A_{02}):
    - controls, if the data satisfy the limitation (CC) (A_{020});
    - controls, if the data satisfy the patterns (PC) (A_{021}) and etc.;
  - Outlier detection – abnormal data (A_1):

  - interquartile range (IQR)) (A_{10});
  - by quantiles P01-P99 (A_{11});
  - Local Outlier Factor (LOF)) (A_{12});
  - detects outliers by means of Zscore, as the function of the median and average absolute deviation (ZSB) (A_{13}) etc.;
- Normalisation (A_2):
  - maximum scaling (MM) (A_{20});
  - decimal scaling (DS) (A_{21}) etc;
- Reduction of the distribution law to normal (A_3):
  - logarithmation (A_{30});
  - square-root generation (A_{31});
  - calculation of the reciprocal values (A_{32});
  - Box-Cox power transformation (A_{33});
- Stationarity provision: n*th* order differences taking (A_4).

Certain researchers, for instance, in the study [6] refer Feature engineering to the pre-processing, but in the given study we suggest to refer this stage to the model identification stage and concentrate directly on the stage pre-processing operations for the model, as FE stage has some specific peculiarities and needs separate consideration.

Formalization of the machine learning with reinforcement problem

As it is known, method of machine learning with reinforcement is based on the formalization of all the operations on the base of Markov's process. One of the main properties of this process is that the following state depends only on the state in the previous moment of time and selected action. Game algorithm is used, when at each «move» or step of the algorithm from this or that admissible set of action variants, using certain «RL-policy» π one action of the system transfer from the i^{th} state S_i into the next state S_{i+1} is chosen and reward R which is achieved is calculated and verified if the final reward is achieved, for which, as a rule, far greater reward is provided (except the cases, when optimal variant is far longer «game»). In case of selection of the bad or inadmissible variant the negative value of the «reward» is established to stimulate the algorithm not to choose it any more. The task is the identification of the sequence of the admissible operations with certain parameters and setting the possibilities for the selection of the «move» variants (directions on the conventional «board» for the «game») of RL-policy π [6 – 8].

The important fact is that the application of RL-model must take place in the conditions as much as possible close to the conditions on which it trained. One of the most valuable possibilities of this technology is that it can operate even if the real data are not available – that is, generate random sequences according to the preset rules and requirements and train itself. And when the real data arrive, it is adapted to these data and corrects its first priority settings. That is why, its name is «learning with reinforcement».

Objective of the research is searching of the optimal pipeline of the operations C, which will provide maximum of the reward by the results of the work of the reinforcement learning method and its application for the obtaining maximally efficiently processed data. Objective function of this method at time t at the set epoch (algorithm iteration) is function G_t [10]:

$$G_t = \sum_{k=0}^{\infty} \gamma^k R_{t+k+1} \,, \tag{1}$$

where R_t – is the function of the reward at the moment of time t, γ – is the coefficient of discounting in the range [0, 1], it determines in how many epochs the impact of the previous values should be «forgotten». If the model for which the preprocessing is performed and the method of its identification are already known, then, as the alternative, in (1) the reward can be taken into account (as a rule, accuracy) only at the last step, but not – average or total reward at all the elements of the pipeline.

From the formula (1) function $V(s)$ of the state value s is obtained for the evaluation of the expected state S_t at the moment of time t [10]:

$$V(s) = E[G_t | S_t = s] = E[\sum_{k=0}^{\infty} \gamma^k R_{t+k+1} | S_t = s], \tag{2}$$

where E – is the operation of taking mathematical expectation.

In the same way, the formalization of the function «state-action» $Q(s,a)$ is realized [10]:

$$Q(s,a) = E[G_t | S_t = s, A_t = a] = E[\sum_{k=0}^{\infty} \gamma^k R_{t+k+1} | S_t = s, A_t = a], \tag{3}$$

where A_t – is the selected action at the moment of time t, on its choice RL-policy of π method on the current epoch influences:

$$A_t = \pi(S_t), \tag{4}$$

which after each epoch or certain number of such epochs is updated to π'. There exist various types of updating of RL-policy π', for instance, in the study [10] 6 methods are presented , they can be generalized by the following expression:

$$\pi' = \mathrm{F}(\mathrm{G}, \mathrm{V(S)}, \mathrm{Q(S,A)}), \tag{5}$$

i. e., for the updating the variant can be used, that provides maximum total reward G (this is the most widely spread variant) but other function of the reward can be where values of the functions $V(S)$ and $Q(S,A)$, calculated for various functions of the state and actions, selected at different stages, by the expression (2) and (3), correspondingly, may be used.

Scheme of RL-method, taking into account (1 – 5) is shown in Fig. 1.

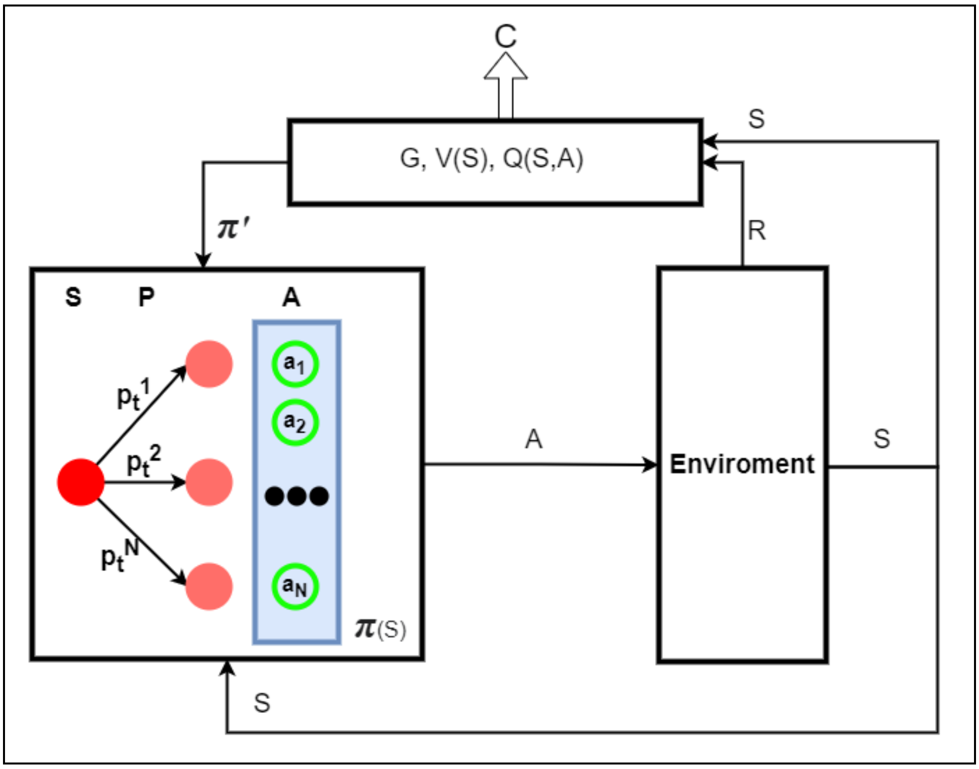

Fig. 1. Scheme of the machine learning with reinforcement method

One of the key components of RL-methods is the algorithm of the reward determination – this provides the convergence, optimality and speed of the result obtaining. Conventionally, any fixed number are chosen, for instance: 1 for the correct intermediate move, -1 or -10 for incorrect, 100 – for final, which ends the game. Another variant – certain simple function, which is the same for all the moves. This simplifies the convergence of the algorithm, but greatly decreases its functionality.

Authors tried to use for the computation of the intermediate reward the following Metrix R_i:

$$R_i = \frac{1}{S_{AUCi}}, \tag{6}$$

where S_{AUCi} – is dimensionless value, that shows how many times the percent of the area under the curve decreased after the i^{th} preprocessing operation («AUC» – «Area Under Curve» – name, used by Python-team library sklearn).

But the experiments with modeling of the atmospheric air quality in the city of Vinnytsia showed (Fig. 2), that for certain indices, values of which are greater than the number e = 2.718…, for instance, after the logarithmation (operation «Ln» in Fig. 2), the values of the reward «R3» can reach great values (Fig. 2a). This considerably complicates the determination of the total reward, as for the correctness of the pipelines comparison it is necessary to calculate the rewards weights, obtained at different moves. But in general case, the pipelines may have various operation sequences and it will be difficult to compare them.

Psmall	R1	Pbig	R2	Ln	R3
0.03	1.03	0.98	1.02	1	8545.04
0.03	1.03	0.98	1.02	1	8545.04
0.03	1.03	0.98	1.02	1	8545.04
0.03	1.03	0.98	1.02	1	8545.04
0.03	1.03	0.98	1.02	1	8545.04
0.03	1.03	0.98	1.02	1	8545.04
0.02	1.02	0.98	1.02	1	8543.47
0.02	1.02	0.98	1.02	1	8543.47
0.02	1.02	0.98	1.02	1	8543.47
0.03	1.03	0.98	1.02	1	8545.04
0.03	1.03	0.98	1.02	1	8545.04
0.03	1.03	0.98	1.02	1	8545.04

a)

Psmall	R1	Pbig	R2	Ln	R3
0.02	1.0	0.97	1.04	1	1.06
0.02	1.0	0.97	1.04	1	1.06
0.02	1.0	0.97	1.04	1	1.06
0.02	1.0	0.97	1.04	1	1.06
0.02	1.0	0.97	1.04	1	1.06
0.02	1.0	0.97	1.04	1	1.06
0.02	1.0	0.97	1.04	1	1.06
0.02	1.0	0.97	1.04	1	1.06
0.02	1.0	0.97	1.04	1	1.06
0.02	1.0	0.98	1.03	1	1.06
0.02	1.0	0.98	1.03	1	1.06
0.02	1.0	0.98	1.03	1	1.06

b)

Fig. 2. Results of the application of the preprocessing operations to the averaged for every 4 hours data of the atmospheric air quality in point № 1315 of the EcoCity network, located on the territory of VNTU, measured during 11.11.2021 – 11.02.2022 : a) ozone concentration O_3, b) concentration of nitrogen oxide NO_2

That is why, we suggested not to use the intermediate reward by the hypothetic formulas but immediately apply the model of machine learning with the set matrix. For this purpose the best variant is to use certain simplified model, rather rapidly calculated, but not models, calculated at numerous graphic processors (GPU) during several hours or days. Besides, it is better to take metrices, the best value of which is 1 and smaller values are worse, for instance in the library sklearn for regression problems it is – «r2_score», and for the classification problems – it is «accuracy_score», but other variants are also possible. As a reward not the value itself of such matrix can be taken, but its square or cube to reinforce the weight of values, closer to 1.

There exist several methods of RL-model optimization [10 – 14]: Q-learning, DQN, PPO, A2C and their variations. In the simplest and thus, most widely used method, Q-learning method preset table of the rewards for all the combinations of states and actions is used. As a rule, the task is greatly simplified to minimize the size of such table. Bright example of the problem solution, applying this method is shown in the work [8] (Fig. 3). The drawbacks are fixed matrix of the rewards, too small dimensionality of such matrix, from which the solution will be performed during limited time and that the process of the optimal operations choice is formed from the end, i. e., the desired model of the machine learning is set and preprocessing methods are selected for this model, such procedure does not always give the correct effect, taking into account that it is better to arrange (tune) the models of the machine learning to data but not arrange data to the models.

DQN – is the method, based on the identification of RL-policy in the form of the deep neural network, applying the principles of the ordinary Q-learning method. As the agent of Q-learning can not assess the states it did not see, 2D grid of the rewards for each combination of the states and actions was replaced by the neural network. That is, DQN method uses neural network for the assessment of Q-function values. Data are sent at the input of the network, at the output the value of Q for each possible action is obtained [8, 13]. Main drawback of such method of reinforcement learning realization is that the discrete space of actions, even at small dimension of the space of action, increases exponentially, thus, it is very difficult to obtain the convergence to the optimal solution during the limited time.

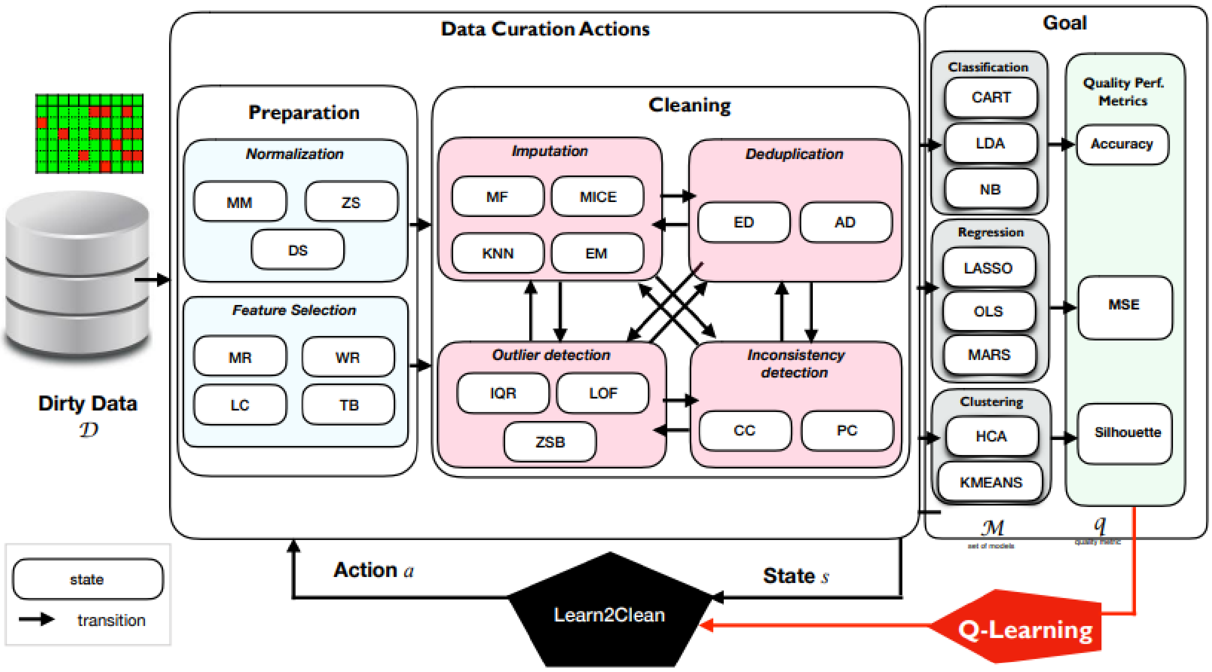

Fig. 3. Algorithm of the method operation, using Q-learning [8]

PPO – is a method with quasi-optimal policy, it is aimed at improvement of the stability of learning policy, limiting changes taking place during each epoch of learning by means of avoiding very critical updating of the policy [10, 14].

Unlike DQN and PPO, A2C – is a method «actor-critic», it has hybrid architecture, that combines methods, based both on the actions and policy change, in particular:

- «actor» controls the behavior of the agent (method on the base of the policy);
- «critic» determines how efficient was this or that operation (method of the base of the actions).

In the process of the learning the «actor» studies the best action, using the responses of the «critics» (instead of using the reward directly). At the same time, «critic» studies the function of value from the rewards, to the able to criticize the «actor». In general these agents can process both discrete and continuous spaces of actions. As a rule, the idea of the method is the accumulation of the chosen routes (sequences, pipelines, combinations) of the probably optimal actions and their periodic analysis in order to correct policy for the improvement of the selection rules of these optimal actions. It is suggested to use this method for the solution of the set task but with certain improvements.

Optimization of the known reinforcement methods for the construction of the pipeline of the data preprocessing means

For the improvement of the accuracy and speed of A2C method or its modifications application for the solution of the set task it is suggested, first, to allocate variable and non-variable links of the general pipeline Ω of the preprocessing operations:

$$\Omega = \beta(C(\alpha(D))), \tag{7}$$

where α – is unchanged link of the serial operations, the expediency of their application is verified first of all for the set data *D* by the simple serial algorithm, *C* – is changeable link with optimal pipeline of the operations, used to the result of the application of the operation α and is identified using the methods of reinforcement learning with, β – unchangeable link of the serial operations, expediency of their application is verified the last for the result of the operation *A* application by the simple serial algorithm. Criterion of changeability and stationarity of the links is their length. If in α and β it is fixed, then, in general case, the link of the pipeline *A* can be of different length.

As it is known, it is worth distinguishing operations, expediency of which is rather easy to determine by means of a simple algorithm and operations, for the determination of the expediency, sequence and parameters of which RL-algorithms are needed. Besides, different models require different approaches regarding data preprocessing for their application.

In particular:

- For all the models data duplicates must be determined and eliminated ;
- all the models must be checked for the availability and elimination from the datasets the erroneous and noncorresponding data;
- models of the library «Science Toolkit for Machine Learning» – abbreviated «scikit-learn» or «sklearn»), as a rule, require the obligatory standardization of the root-mean-square data deviation with centering around average value of data (operation «StandardScaler»);
- models of the time series ARIMA, as a rule, require the data with the same time step without any blanks, thus, the obligatory verification of the data for the absence of the omitted data and their interpolation, in other models (Facebook Prophet and others) elimination of the data (rows of the table) ,where the omitted ones may be present, is often practiced, as any interpolation distorts the original data and can introduce the erroneous regularities;
- if nonspecific for time series models of the machine learning (solution trees or nonrecurrent neural networks) are applied for the prediction of periodic time series it is necessary to identify and eliminate recurrent additive or multiplexing trend (specific models of ARIMA, Facebook Prophet, etc. have the built-in operations for this purpose);
- data scaling is needed for the graphs construction in one coordinates system, and this does not influence the operation of the models of machine learning.

Operation for duplicates elimination must be the first, elimination of the erroneous and noncorresponding data – must be second, interpolation, if necessary – the last but one, standardization, if needed, must be the last operation, to have the possibility of execution the convenient inverse transformation, if it is needed:

$$\alpha = A_{02}(A_{01}(D)),\ \beta = A_4(A_{00}(C(\alpha(D))) \tag{8}$$

Between these operations more complex operations of the variable link *C* must be performed, the sequence, availability and parameters of which are to be identified, applying RL-methods both for random values and time series:

1. Filtration of the anomalies in greater values, for instance, by means of the filter selection according to the values of quantiles P01-P25 or P75-P100 (26 variants, including variant P100, which means the lack of filtration).

2. Filtration of the anomalies in smaller values, for instance, by means of the filter selection according to the values of quantiles P00-P26 (26 variants, including variant P00, which means the lack of filtration).

3. Reduction of the distribution law to normal (logarithmation, taking inverse values or root, etc., – different variants are possible).

4. For time series: provision of the stationarity of the time series by means of taking difference of the n^{th} order, where $n = 0, 1, 2, 3$ (greater values of n, as a rule, are not used), and $n = 0$ corresponds to the lack of expediency to use this transformation (4 variants).

5.For time series: non-linear transformations of the trend, including, periodic and its extraction from the series of values (different number of variants is possible).

Other operations are possible. But first 3, and especially, first 2, are the most popular and efficient.

Such optimization of the algorithm, taking into account (8) enables to optimize the speed and involves more operations at each step from the set of their admissible variants.

Another optimization can be achieved if previous experience and the results of the intelligence data analysis are taken into account. One of the conventional approaches to the operation of RL-methods is that there is no preliminary information about the input data. But the availability of such information enables to accelerate considerably the convergence of the algorithm to the optimal solution due to the setting of the probabilities in policy π not randomly but according to certain algorithm:

1. Anomalies filtration in greater values («from the top»). It is suggested to select the probabilities proportionally to the values of quantiles, scaled in the range [0, 1] in such a way that maximum values (as a rule, this is maximum, that conventionally can be denoted as P100) are selected more often, and P75, at which most of the data are lost – seldom. Another variant – set probabilities by Gauss curve so that P75 and maximum were selected seldom and P87 and neighboring – more often.
2. Anomalies filtration in smaller value («from below»). Similarly to the point 1, but in the inversed manner: probabilities should be selected inversely proportional to the values of the quantiles, scaled in the range [0, 1], so that minimal values (as a rule, it is minimum that can be conventionally denoted as P00) were selected more often, and P25, at which greater part of data are lost – seldom. Another variant – set the probabilities by the Gauss curve, so that P25 and minimum were selected seldom and P12 and neighboring – more often.
3. To provide series stationarity it is necessary to take into account that the optimal is the variant that provide p-value (or p) less than 0.05, but as close as possible to this value. That is, if the verification of the hypothesis that the value of the series are distributed not according to the normal law gives p-value, that is equal 0.1, and regarding the first difference – 0.04, the second – 0.02, it means that it is worth taking the first difference but not the second, as greater order means greater loss of the information. That is why, it is expedient to calculate J_s criterion of p-value closeness to 0.05 for each variant:

$$J_s = \begin{cases} e^{-(p-0.5)^2}, p \le 0.5, \\ 0, p > 0.5, \end{cases} \quad (9)$$

or take the probabilities $1/n$, $n = 1, 2, 3$, that is, inversely proportional to their order, scaled in the range [0, 1], and for $n = 0$ take for instance, the value 2, as to treat the results of modeling for the series itself is more convenient than for its difference.

In the same way all the probabilities of the policy π for each set of actions at each move (iteration) can be arranged. It is recommended to set the largest frequency or probability for each variant, when the operation is not used, as any operation makes changes in the primary dataset, but it is not desirable.

It is also expedient after the operation of the algorithm at certain intervals to update the a priori frequencies (probabilities) of the policy π for each set of actions by the results of modeling – a posterior.

Conclusions

The problem of the synthesis of the optimal pipeline of the preprocessing operations for the set models is expedient to perform, using the methods and technologies of the Reinforcement Learning – RL, as other methods do not provide the convergence to the optimal variant, if relatively large number of preprocessing operations with different sequence and parameters are taken into account.

Formalized problem setting has been performed. The selection of the method «A2C» is substantiated as the best one for the solution of this problem with simultaneous optimization of actions and states in different ways. It is noted that RL-methods also have problems with the convergence and operation speed on conditions of the great amount of the possible operations. That is why, several improvements have been suggested. It is worth performing the decomposition of the operations pipeline into the sequence of the unchangeable by the length links and one middle link of the potentially changeable length. Simple algorithms for the unchangeable links are suggested, for the changeable link it is expedient to use complex RL-algorithms – this enables to reduce considerably the dimensionality of the problem. Besides, the approaches to the selection of the initial policy or algorithm for the selection of actions at the first epoch (iteration) of the algorithm, taking into account the statistical indices and other regularities of the input data have been suggested. All this enables to increase the chances to find really optimal solution during the acceptable time.

The suggested improvement of the method with the reinforcement of the synthesis of optimal pipeline of the operations can be used not only for preprocessing operations in machine learning problems but for other problems with similar data formalization and problem set up.

REFERENCES

1. Mokin V. B. Analysis of the experimental data, needed for the synthesis of the mathematical model for the prediction of the allergenic mushrooms Alternaria spores spreading / V. B. Mokin, V. V. Rodinkova, M. V. Dratovanyi // Bulletin of Vinnytsia Polytechnic Institute. – 2019. – № 3. – P. 50 – 58. (Ukr).

2. Environmental Factors Which Increase Alternaria spores in Central Ukraine / V. Rodinkova, V. Mokin, O. Bilous, [et al.] // Journal of Allergy and Clinical Immunology. – 2018. – № 141 (2). – P. AB30.

3. Prediction of the coronavirus waves on the base of the cognitive map of the interregional impact / V. B. Mokin, M. V. Dratovanyi, A. V. Losenko [et al.] // Information technologies and computer engineering. – 2021. – № 3. – P. 86 – 94. (Ukr).

4. Intelligent technology for the analysis and prediction of the prices for the used motor vehicles / V. B. Mokin, A. V. Losenko, M. V. Dratovanyi [et al.] // Bulletin of Vinnytsia Polytechnic Institute. – 2019. – № 6. – P. 62 – 72. (Ukr).

5. Automatically Debugging Auto ML Pipelines Using Maro : ML Automated Remediation Oracle (Extended Version). [Електронний ресурс] / J. Dolby, J. Tsay, M. Hirzel. – 2022. – Режим доступу : https://arxiv.org/pdf/2205.01311.pdf.

6. Reinforcement learning for data cleaning and data preparation [Electronic resource] / L. Berti-Equille // Proc. ACM SIGMOD Int. Conf. Manag. – 2019. – Access mode : https://laureberti.github.io/website/pub/HILDA2019_Laure.pdf.

7. Automated Data Preprocessing for Machine Learning Based Analyses [Electronic resource] / A. Paranjape, P. Katta, M. Ohlenforst // COLLA 2022 : The Twelfth International Conference on Advanced Collaborative Networks, Systems and Applications. – 2022. – Access mode : https://www.researchgate.net/profile/Akshay-Paranjape/publication/361026018_Automated_Data_Preprocessing_for_Machine_Learning_Based_Analyses/links/629

8a13c6886635d5cb84dee/Automated-Data-Preprocessing-for-Machine-Learning-Based-Analyses.pdf.

8. Scenario-based automated data preprocessing to predict severity of construction accidents [Electronic resource] / K. Koc, A. P. Gurgun // Automation in Construction. – 2022. – Volume 140. – P. 104351. – Access mode : https://www.sciencedirect.com/science/article/abs/pii/S0926580522002242?via%3Dihub.

9. Mokin B. I. Mathematic methods of the dynamic systems identification : manual / B. I. Mokin, V. B. Mokin, O. B. Mokin. – Vinnytsia : VNTU, 2010. – 260 p. (Ukr).

10. Benchmark study of reinforcement learning in controlling and optimizing batch processes [Electronic resource] / W. Zhu, I. Castillo, Z. Wang [et al.] // Journal of Advanced Manufacturing and Processing. – 2022. – № 4 (2). – Access mode : https://doi.org/10.1002/amp2.10113.

11. Reinforcement learning for data preparation with active reward learning [Electronic resource] / L. Berti-Equille // In Internet Science : 6th International Conference, INSCI 2019, Perpignan, France, December 2 – 5, 2019. – Proceedings 6. – P. 121 – 132. – Springer International Publishing. – Access mode : https://doi.org/10.1007/978-3-030-34770-3_10.

12. Papoulis A. Brownian Movement and Markoff Processes / A. Papoulis // Ch. 15 in Probability, Random Variables, and Stochastic Processes, 2nd ed. – New York: McGraw-Hill, – 1984. – P. 515 – 553,

13. Competitive reinforcement learning in Atari games / M. McKenzie, P. Loxley, W. Billingsley [et al.] // Australasian Joint Conference on Artificial Intelligence. – 2017. – P. 14 – 26.

14. Learning from learners : Adapting reinforcement learning agents to be competitive in a card game / P. Barros, A. Tanevska, A. Sciutti // In 2020 25th International Conference on Pattern Recognition (ICPR). – 2021. – P. 2716 – 2723.

Editorial office received the paper 24.12.2022.
The paper was reviewed 28.12.2022.

Dratovanyi Mykhailo – Assistant with the Department of system analysis and information technologies, e-mail: mishadratovany@gmail.com.

Mokin Vitaliy – Dr. Sc. (Eng.), Professor, Head of the Department of system analysis and information technologies, e-mail: vbmokin@gmail.com.

Vinnytsia National Technical University.