UDC 004.9

Reinforced machine learning methods for testing quality of cyber threat prediction results

A. B. Kachynskyi[1], N. A. Tsebrinska[2]

[1] *National Technical University of Ukraine «Igor Sikorsky Kyiv Polytechnic Institute», Institute of Physics and Technology*

[2] *National Technical University of Ukraine «Igor Sikorsky Kyiv Polytechnic Institute», Institute of Physics and Technology*

Abstract

The article considered on machine learning methods with reinforcement to make decisions about evaluating the quality of a mathematical prediction model. Given the problems of cybersecurity specificity A/B testing algorithms, analysis of variance (ANOVA), as well as multi-armed bandit are presented. Features of their practical implementation are taken into account: data type and distribution function, sample size, knowledge about the dispersion of the general population, dependence and independence of observations. The cybersecurity problems solved with the help of these algorithms are discussed and the methods of their solution are suggested.

Keywords: Reinforced machine learning, hypothesis testing, A/B testing, analysis of variance, multi-armed bandit

Formulation of the problem

Now, in the regional scientific research that has solved the problem of safety, machine learning methods have begun to show high efficiency. Currently, they provide a answer to the following important cybersecurity issues [10]:

- Does every file transmitted over the network contain a malware?
- Does one of the employees use a compromised password when trying to log in?
- Does every received e-mail contain a phishing attempt?
- Is there a DOS attack for each request to the network server?
- Is each outbound request sent to bot that invokes its server to intercept operational controls?

All of the above tasks are classification problems, that is, decision-making tasks to determine the nature of the phenomenon being observed [1] . Thus, the basic principle of the work tasks of providing information and cyber security with the help of machine learning systems is the following: to classify all events on the network as inadmissible (harmful) or admissible (correct). Typically, this is achieved by logging binary logs, logging attempts, emails received, incoming and outgoing requests, etc., and searching for pre-recorded pattern data that indicate malicious attacks. The next step is to encode these templates as an algorithm, that is, a function that accepts the input data that needs to be classified and get a binary response: "harmful" (data) or "harmless" (data). Thus, if one uses the algorithm of scanning chronological data from the past and finds the best classification rule according to some mathematical definition of "best", then this process is called machine learning with a teacher in information and cyber security.

Analysis of research and literary sources. During the research of the features functionality and development of machine learning systems cybersecurity researchers have been found the process of deterioration the quality of their work during prolonged operation. The analysis of system-wide regularities that characterize the fundamental features of the structure, operation and development of complex systems will help more deeply understand causes of this phenomenon [7].

The communicative macro parameter is the cornerstone of the theoretical foundations of general security theory. He claims that system cannot exist in isolation from another systems, it is connected with communications with its environment, which is complex and heterogeneous. Over time, its external influences can change the landscape of information and cyber threats.

The equiprobability macro parameter characterizes the system's marginal capabilities. According to this system-wide regularity, an equifinal state is a state opposed to the equilibrium state in closed systems that determined by initial conditions. Equifinal state is reached regardless of time and initial conditions and is determined solely by the parameters of the system itself. For machine learning systems with a cyber security teacher, this means that in the future, the organization's management may have to change the concept of the organization's security policy.

Macro parameter of "requisite variety" (W. Ashby Law). Using this macro parameter for improvement functionality of cyber security machine learning systems will help to identify the reason of their deficiencies and ways to overcome them. Given the above: the variability of the landscape of information or cyber threats and changing the security policy of our organization, we can concluded that one of the main reasons of degradation machine learning systems is changing the quality of input data and the size of the characteristic space.

Macro parameter of historicity: life cycle. Although the variability of system parameters is obvious, in recent times the general security theory in the design and management of complex security systems has been increasingly paying attention to common factors of historicity[9]. Moreover, this common factors can be taken into account not only for the passive fixation of mathematical model degradation of classification dangerous phenomena of virtual space. The lifecycle model, considering the principles of changes in machine learning systems over time, can be used to improve the quality of their work and prevent degradation.

Formulating the goals of the article. When machine learning security systems become sensitive to data quality, the first step to ensuring flexibility of system is detecting the actual fact of degradation of the mathematical model [15]. Feedback cycles are an appropriate way not only to detect deterioration in the model's performance, but also to improve it continuously.As a result, machine learning systems can be considered as the most suitable systems for detecting anomalies and cyberattacks with an integrated feedback loop.

Reinforcement learning is a machine learning method when the model trains using dynamic programming techniques, Monte Carlo, time differences, and a feedback loop in the form of rewards (positive feedback signal) and punishment (negative feedback signal).

Training is complicated by the fact that the promotion is rarely given and only after completing the entire sequence of actions. Promotion determines the purpose of the task and must be obtained if we want to learn. The agent learns in the best sequences of action, leading to the task being solved. The best course of action means that it is possible to achieve maximum encouragement as soon as possible[13].

One of machine learning method with reinforcement is the method of active learning, which include the partial involvement of the teacher. In this method, the training classifier model should select the data items for which it is less likely to perform predictions and invite experts to assign labels to these unreliable data. Using feedback, experts assign the correct data to the label. In practice, this task should be performed by experienced security analysts with sufficient qualifications. After that, the algorithm uses them to train and improve the mathematical model. Active learning is a useful way to predict threats. In the field of information and cyber security, there is always a lack of sufficient data with the correct labels.

Thus, in the machine learning method of reinforcement and its variant - method of active learning there is no external process of providing data for training. Machine learning is able to efficiently process large amounts of data to detect patterns and anomalies, while the agent is actively generating data, experimenting with the environment and receiving feedback in the form of rewards. Then he uses feedback to correct his knowledge in order to learn how to implement the actions that lead to the greatest encouragement and adaptation of the mathematical model to changes in the environment.

Presentation of the main material

Currently, there are several strategies for implementing machine learning with reinforcement for predicting and reversing cyberattacks: A/B testing, ANOVA, k-arm bandit. As a feedback tool for mathematical model correction, this type of mathematical training not only takes into account changes in the security environment, but also assigns labels with unreliable data, draws reasonable conclusions about the dimension of characteristic space, and can also reveal common causes of degradation of the security system.

A/B-testing

This type of machine training was not offered by security professionals. In practice, the first time A/B testing was used by Greg Lindgren, one of the founders of Amazon, to develop commercial referral systems. Subsequently, technicians who had a high level of knowledge in mathematical statistics adapted this technology to evaluate the quality of predictive mathematical models [2; 12].

However, in order to successfully combat cyber threats, it is not enough to be able to use the standard approach: deep knowledge of the mathematical foundations of this method regarding to the type of data, sample size, correction coefficients and etc. is required. Without this knowledge, it is impossible to apply the concept of machine learning with reinforcement to solve the real problems of cybersecurity, which in this article is considered in the context of analytical forecasting.

The standard A/B testing procedure is implemented as an experiment with two groups to determine which one is better with a particular metric. The current version of the object is called the control version, while the modified version is called the calculated version. In the experiment, both versions participate simultaneously.

Using machine learning systems to solve cyber security problems always necessary to test new models (calculated version) with in-service models (control version) using A/B testing. Performing this experiment, it is necessary to have a well-defined metric (for example, the number of spam emails, malicious files, etc.) [15]. These metrics are measured by user feedback or by sampling data and assigning them to them.

For machine learning systems, A/B testing is really important, since the gradual updating of long-running models (for example, by re-training models, adjusting data, etc.) may not produce the expected results. Experimenting with new models and empirically determined conditions of achieving best performance adds machine learning systems the flexibility they need to adapt to changes in data and algorithms related to cybersecurity.

The algorithm for constructing the test criteria for A/B-testing hypothesis of machine learning systems for cyber security will be as follows.

Step 1. Formation of a verbal security model of information situation.

The purpose of this step is to enable the researcher to get acquainted with the general view of the set of data to which this statistical criterion should be applied.

A security information situation means a certain level of uncertainty about the information environment location in one of its possible states throughout A/B testing.

Comparative analysis of two groups (A/B-test) is a necessary element of mathematical statistics and is covered in many fundamental works [3; 4; 11]. However, this does not apply to A/B testing thematic materials in the context of safety aspects. Therefore, it is necessary to be cautious about the advice covered in popular works and its application [2; 10; 12;]. With regard to the practical application of A/B testing to design experiments and statistical decision-making procedures in a secure environment, the following information situations are possible: continuous data (large sample), continuous data (small sample), continuous data (temporal pairwise comparisons,) discrete data.

Step 2. Formation statistical model of information situation.

This is the key point of the algorithm. The random nature of the statistics is assumed that include conditions for the distribution (type of data), independence (or dependence) of observations. They must necessarily meet the real conditions. On the other hand, you can often neglect slight deviations from normalcy.

If the probabilistic model on which the criterion is based does not match the data, then it is possible that the conclusions that can be drawn from the proposed criterion will also be erroneous.

Step 3. Establish a null hypothesis H_0 and an alternative hypothesis H_1.

In case of A/B-testing, the null hypothesis is considered when we are testing the hypothesis of the significance of the difference between the sample averages:

$$H_0 \ : \ \mu_1 \ = \ \mu_2$$

Where μ_1 and μ_2 are general averages for groups A and B, respectively. For two-sided verification, the alternative hypothesis is:

$$H_0 \ : \ \mu_1 \ \neq \ \mu_2$$

In one-sided verification, the following cases are possible:

$$H_1 \ : \ \mu_1 \ > \ \mu_2,$$
$$H_1 \ : \ \mu_1 \ < \ \mu_2.$$

Step 4. Define type I and type II errors. Type I error (denoted by α) corresponds to the wrong decision, which is made without accepting the null hypothesis at time when it is valid for the general population. Type II error (denoted β) is made when the null hypothesis is accepted, when it is in fact incorrect for the general population. The probability of a second-order error β depends on an alternative hypothesis H_1.

In the theory of mathematical statistics, errors of the type II have traditionally received less attention: historically, this error was considered to be less serious than the type I error. For type II error, thresholds of 0.1 or 0.2 were assumed, which meant 10% and 20%, respectively, of the probability of mistakenly assuming a hypothesis when it was incorrect.

In recent years, security theory has become more focused on the study of the required power level of the criterion (1 - β). This is first and foremost related to the definition of acceptable risk [9]. In addition, the calculations of the power of the criterion, and thus the errors of the second kind, play an important role in the planning of safety studies, especially in determining the sample size required to achieve sufficient power of the criterion. Thus, solving security problems, is important to choose the most powerful criterion that minimizes α and β.

Step 5: Search for critical areas. Preferably a 5% significance level is offered for all criteria. It is not difficult to move to another level of significance for the criterion α : 0< α <1 (for example, 0.1; 0.05; 0, 01). This level indicates the exact value of the first-order error probability, if the hypothesis is indeed correct. In other words, the level of significance means the magnitude of the risk of making a first-class mistake. The lower the level of significance, the lower the likelihood of first-order error. However, the lower the level of significance, the greater the likelihood of a second-order error if the hypothesis is false.

Step 6. The values of the criterion statistics are calculated based on the statistics. Setting up a computing procedure.

Step 7. Compare the values obtained performing step 6 with the values of the critical values of step 5. Based on this comparison, a decision is made to accept or reject the null hypothesis H_0.

Step 8. Comments. Various comments are made regarding calculations, alternative criteria, null hypotheses, independence of observations, and etc. Their main goal is to find out if the observed effect can cause accidental changes.

Information situation 1.

Two normal samples are regarded: the calculated version of the cybersecurity machine learning model (sample A) and the model control sample (sample B) with parameters $N(\ \mu_1,\ \sigma_1)$ and $N(\ \mu_2,\ \sigma_2)$ and the number of independent observations n_1 and n_2. It is necessary to determine which of these models is better: less erroneous cyber-predictions are assumed.

In addition to the above, the following conditions must be met in this A/B test situation:

- large sample: $n_1 \geq 30$ and $n_2 \geq 30$;
- known dispersion.

If dispersion of statistical population is known, then the criterion relations will be:

$$Z \ = \ \frac{(\overline{x}_1 - \overline{x}_2) - (\mu_1 - \mu_2)}{\sqrt{\dfrac{\sigma^2{}_1}{n_1} + \dfrac{\sigma^2{}_2}{n_2}}} = \frac{\overline{x}_1 - \overline{x}_2}{\sqrt{\dfrac{\sigma^2{}_1}{n_1} + \dfrac{\sigma^2{}_2}{n_2}}}$$

and if dispersion of statistical population is unknown:

$$Z = \frac{\overline{x}_1 - \overline{x}_2}{\sqrt{\frac{s^2{}_1}{n_1} + \frac{s^2{}_2}{n_2}}}$$

$\overline{x_1}$ and s_1 – average value and standard deviation of sample A with size n_1, $\overline{x_2}$ and s_2 – average value and standard deviation of sample B with size n_2

The rejection or acceptance of the null hypothesis is carried out according to the results of calculations obtained by the standard procedure of hypothesis testing.

Information situation 2.

This situation can occur when the mathematical model classification of dangerous phenomena of virtual space with long life cycle is degraded, which is confirmed by statistics for the previous periods of time. In this case, the procedure for testing the hypotheses regarding the difference of the general averages is similar to the previous case of large samples. However, there are differences: large sample: $n_1 \geq 30$ and $n_2 \geq 30$;

- observations are continuous;
- dispersion σ_1 and σ_2 are different;
- small sample: $n_1 < 30$ and $n_2 < 30$;
- critical value based on Student's t-distribution.

The evaluation criterion is ratio:

$$t = \frac{\overline{x}_1 - \overline{x}_2}{\sqrt{\frac{s^2{}_1}{n_1} + \frac{s^2{}_2}{n_2}}}$$

This ratio has an approximate t-distribution for the corresponding number of degrees of freedom df. A special case is the equality of dispersion of general groups: $\sigma_1^2 = \sigma_2^2$. In practice, dispersions are considered equal unless there are strong arguments to the contrary. Consider the problem of hypothesis testing under the following conditions:

- sample size does not exceed 30;
- the samples are taken from two different statistical population independently of each other;
- both statistical populations are approximately normal;
- dispersions of statistical population are equal;

$$H_0 : \mu_1 = \mu_2,$$
$$H_1 : \mu_1 \neq \mu_2.$$

Then the value is a criterion [14]

$$t = \frac{\overline{x}_1 - \overline{x}_2}{\sqrt{\frac{(n_1-1)s^2{}_1 + (n_2-1)s^2{}_2}{n_1+n_2-2} + \left(\frac{1}{n_1} + \frac{1}{n_2}\right)}}$$

For the significance level α, the value is compared with the value $t_{\alpha/2;df}$, where df $= n_1 + n_2 - 2$. If , then the null hypothesis cannot be rejected with the significance level α. When $|t| \geq t_{\alpha/2;n_1+n_2-2}$ the difference of averages falls into the critical region and the null hypothesis H_0 is rejected at the significance level α. For one-sided inspections, the relevant critical limits are considered $\pm$ $t_{\alpha;n_1+n_2-2}$

Information situation 3: discrete data

As noted, scanning algorithms often provide data in the form of categorical or discrete metrics to help make predictions about cybersecurity machine learning forecasts.

There are two statistical population represented by the number of fates to be compared. Select all possible samples from population 1 by volume n_1, and from population 2 samples with size n_2. The value of the general fate for the population: $1 - p_1$, and for population: $1 - p_2$. While at the same time the conditions are fulfilled: $n_1p_1 > 5$ and $n_1(1-p_1) > 5$, $n_2p_2 > 5$ and $n_2(1-p_2) > 5$, then as the author notes [14], the distribution of differences of sample shares will have a normal distribution function. In this case, the null and alternative hypotheses for the equilateral criterion are written as follows:

$$H_0 : p_1 = p_2;$$
$$H_1 : p_1 \neq p_2.$$

For the significance level α acceptance area of hypothesis H_0 will be conditioned:

$$\left| \frac{\overline{p}_1 - \overline{p}_2}{\sqrt{\widehat{p}(1-\widehat{p})\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}} \right| < Z_{\frac{\alpha}{2}}$$

Accordingly, the critical region is given by inequality:

$$\left| \frac{\overline{p}_1 - \overline{p}_2}{\sqrt{\widehat{p}(1-\widehat{p})\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}} \right| \geq Z_{\frac{\alpha}{2}}$$

Information situation 4:

In previous cases of A/B testing, the necessary conditions for comparing the average of the two samples were their independence and normality of distribution functions. However, using the chronological scan algorithm, there is often a situation that requires pairwise consideration: when pair of security elements of the observation corresponds to the same point in time (for example, to the same provider, server, or even host) .

In this situation, the following conditions must be met:

- dependence of observations (temporal pairwise comparisons);
- data continuity.

There is n observed pairs : $(x_1, y_1), (x_2, y_2), ..., (x_n, y_n)$. Consider the differences as they correspond to each pair:$d_1 = x_1 - y_1, d_2 = x_2 - y_2, ...d_n = x_n - y_n$. Thus the problem of comparing two sets of data is reduced to analyzing one set consisting of differences d_i. For this purpose, calculate the sample standard deviation of the difference [11]:

$$s_d = \sqrt{\frac{\sum_{i=1}^{n} d^2{}_i + \left(\sum_{i=1}^{n} d_i\right)^2 /n}{n-1}} = \sqrt{\frac{\sum_{i=1}^{n}(d_i - \overline{d})^2}{n-1}}$$

where $d = \frac{\sum d}{n}$ - sample mean difference.

Considering that each element in the pair is taken from the normal statistical population, the distribution of sampling differences will also be normal. Suppose we need to test the hypothesis that the mean of this distribution μ_d is D_0 , then the problem of hypothesis testing can be written as follows:

$$H_0 : \mu_d = D_0$$
$$H_1 : \mu_d \neq D_0$$

If the number of observed pairs greater than 30 (n $\geq$ 30), then the standard normal distribution can be used for checking. In the case if observed pairs less than 30 (n <30), we use the Student's t-distribution for checking. The ratio is checking:

$$t_d = \frac{d - D_0}{s_d/\sqrt{n}}$$

which is compared with the values of the t-distribution at df = n - 1 and given significance level α. If $|t_d| > t_{\alpha/2;n-1}$, then the hypothesis H_0 is rejected with the significance level α.One-sided checks are performed similarly: $H_1 : \mu_d < D_0(t_d > t_{\alpha;n-1}$) or $H_1 : \mu_d > D_0(t_d > t_{\alpha;n-1}$)

Analysis of variance (ANOVA)

In many cases, the results of forecasting for audits, debugging and appeals obtained with the help of machine learning methods need to be reaffirmed. So, cyber-incident forecasting systems must be resistant to random disturbances during all decision-making steps.

Suppose that while we are using machine learning systems to solve cyber security problems, it is necessary to check several models of the calculated version with the model of the control version, opposed to A/B testing, it is necessary to compare several groups (for example, A-B-C-D). ANOVA is the statistical procedure that verifies the statistical significance of the difference between groups. It uses the additive property of the random variable variance due to the action of independent factors.

During this process is made decomposition of the total sample variance into components due to independent factors. Each of these components is an estimate of the dispersion of the statistical population. In order to evaluate the effect of the influence of this factor, it is necessary to evaluate the significance of the respective sample variance in comparison with the variance of the reproduction caused by random factors. The significance of the variance estimates is verified using the Fisher test. If the calculated value of the Fisher's test is smaller than the tabulated value, then there is no reason to consider the influence of the investigated factor significant.Depending on the number of dispersion sources one-factor and multi-factor analysis of variance are distinguished. In the analysis of variance data of multi-factor experiment are used the same principles and calculations of variances as in the one-factor experiment. The calculation scheme for testing the null hypothesis is standard and is given in investigation [6; 16].

In case both: A/B testing methods discussed above and the one-factor ANOVA, only one metric is changed (for example, the accuracy of the forecasts) and the remainder is constant. However, in the actual operating conditions of mathematical models, there are a number of factors that influence the successful operation of machine learning systems. This is especially true of cybersecurity, where these systems are in the real-life sector of immediate action.

This approach to solving cybersecurity problems needs a multifactorial study of the effects and interactions of several factors and their impact on the variability of a productive trait. Moreover, each factor was given several gradations [8]. This allows you to study the action of each of them with several gradations of other factors.

Analysis of variance, as practice of its application shown is especially effective in the study of such problems. Moreover, for each of them there are a number of observations that are not used in the study of other factors. This method of study does not determine the interaction of several factors while changing them. It is important that in the analysis of variance each observation serves to simultaneously evaluate all factors and their interaction.

The effect of factors interaction is that part of the general variability caused by different action of one factor at different gradations of another. In a real experiment, often the effect of the joint application of the investigated factors may be higher (synergism) or lower (antagonism) the sum of the effects of the separate application each of them. In the first case there is a positive, in the second - a negative interaction of factors. If the factors do not interact, then the effect of the joint application is equal to the sum of the effects their separate application [5].

Taking into account many factors of the security environment is an integral part of the decision-making process, but more importantly, the effect of reproducing and interpreting the results enables analysts and staff to analyze, evaluate, and fine-tune these systems.

The algorithm of the multi-armed bandit

During A/B testing, the main problem is determining the amount of traffic passing through the billing version A and the amount of traffic passing through the control version B. According to [16], this is one of the variations of the multi-tasking problem, or a k-arm bandit. In this case, it is necessary to maintain a balance between research in order to gain new knowledge and practical use of previously acquired knowledge.

In this way, multi-armed bandit algorithms offer their approach to testing cybersecurity systems, allowing them to more effectively solve the optimization problem and make faster decisions than traditional optimization methods and statistics.

A multi-armed bandit is a hypothetical slot machine with k launching levers, or arms that can be randomly clicked by the player, with each hand receiving different rewards.

Bandit algorithms, which are very popular in web-based testing, allow you to test multiple variants and draw conclusions faster than traditional statistical decision-making methods. They derive their name from the gambling machines known as the one-armed bandits (since they are configured to make as much money as possible from the player). If you imagine an automatic machine that has more than one starting lever or arms, and each of them receives different rewards, then we get a multi-armed bandit, which will give the full name of this algorithm.

The goal is to win as much as possible, namely, to identify the winning lever in as few trials as possible. The problem is that you do not know at what rate winnings are payed - you can only find out after pushing the lever [10; 12].

As noted, the traditional A/B test is related to the data obtained from the experiment according to the defined plan. The main purpose was to answer the question: "Which is better: Option A or Option B?" As soon as we get an answer to this question, the experiment ends and we make a decision based on the results.

However, routing or trafficing a new system there is a need to obtain a large amount of information (for example, the maximum amount of statistics to test). It is also important to avoid the overall degradation of metrics because the performance of the new system A may be worse than the performance of the existing system B.

With this approach there are a number of difficulties [6]:

first of all, the answer may be inconclusive: "effect not proven". Experiment results may indicate that an effect is present, but if it is installed, we can define that the number of observations in the sample is not sufficient to confirm it.

secondly, there may be a situation that requires the results to be used before the experiment is complete.

third, we may change our minds if we receive additional data after the experiment is completed.

There are several approaches to overcome these complications, including two limiting ones. The essence of the first is to say: "It seems that lever A is winning - it is necessary to stop experimenting with other levers and choose A". This approach takes full advantage of the initial information. If lever A is really better, then we get the benefit right from the start. On the other hand, if the B or C levers are actually better, then we lose the opportunity to find it out.

The essence of the second limiting approach is to say: « It all depends on the case, so we will press the levers with equal probability.» This approach gives the greatest chance for other alternatives to prove themselves. At the same time, in its implementation, we can consider variants that are less profitable.

The "bandit" algorithms are also characterized by a hybrid nature - more often to press the lever A, using its obvious advantage, but also to be interested in the levers B and C. We continue to press the levers B and C, but more often we press A. On the other hand, if C begins to work better and A - worse, then we can change the accents by shifting them from A to C. If one of these variants is found to be more successful than A and was hidden during the initial tests by chance, then it now has a chance to prove itself during further testing.

One such algorithm is the epsilon-greedy (ε-algorithm) for the A/B test.

1. Generate a random number between 0 and 1;

2. If this number is between 0 and ε (with $0 < \varepsilon < 1$ and usually quite small), toss a "correct" coin (with probability 50/50), and:

- if the coin drops an eagle, suggest option A;
- if the coin drops a tails, suggest option B;

3. If this number is greater than ε or equal to ε - select the most effective option at this time.

Epsilon (ε) is the only control parameter of the algorithm. If $\varepsilon = 1$, then we end the experiment with a standard A/B test. If $\varepsilon = 0$, then we end with a purely greedy algorithm, preferring the most efficient option.

With regard to such an important issue as avoiding metric degradation during routing and traffic monitoring, we need to use the more sophisticated "Thompson sampling" algorithm here. This sampling procedure (pushes the bandit's starter lever) at each stage of the experiment to maximize the likelihood of choosing the best lever. It is unknown which lever is the best, but as you monitor the payouts with each subsequent recess, you get more and more information. That is, a sample of Thompson, characterized by the amount of traffic for each routing option, is proportional to the probability of getting a better result in the future.

In the Thompson method, a Bayesian approach is used: first, a priori beta distribution of rewards is considered; From the point of view of choosing the right lever, the information after each recess can be updated to better optimize the next recess. This approach is widely used by contextual multi-armed bandit, adding additional factors to the process.

In practice, "bandit" algorithms can handle three or more variants, while making the best choice. As for the traditional statistical testing procedures discussed above, the efficiency of the selection process in bandit algorithms for three or more variants far exceeds them [1]. Table 1 summarizes the research findings, including cybersecurity issues and the features of their practical implementation.

Conclusions

The presence of feedback cycles in machine learning systems can make them highly adaptable to degradation of data quality. But in an unreliable environment, the feedback of the feedback loop in the system is directly impacted - the calculated version of the model can have negative effects, and machine-based security training will affect cyberattacks. Safety systems based on machine learning methods meet the requirements of the input quality.

Studies have shown that A/B testing based on statistical hypothesis checking is possible for any metric and for any statistical criterion. The corresponding probabilistic distributions of sample averages can be of two

Table 1. Reinforced machine learning methods for making decision regarding to the quality assessment of mathematical prediction model

№	Method	The proposed algorithm	Problems solved
1	A / B - testing	Information situation 1	Model quality check for continous metric model (large sample)
		Information situation 2	Model quality check for continous metric model (small sample)
		Information situation 3	Model quality check for discrete metric model
		Information situation 4	Checking quality of the model version for dependent metric
2	Analisis of variance	One-factor analisis	Checking quality of the billing version for multiple levels of one metric
		Multi-factor analisis	Checking quality of the billing version for multiple levels and metrics
3	Multi armed bandit	ε-greedy algorythm, Thomsom method	Determining ammount of traffic going through hte billing and control versions

types: normal distribution and Student's t-distribution. The choice of criterion depends on the sample size and whether or not the standard deviation σ is known.

In actual operating conditions, there are a number of factors that influence the successful operation of machine learning systems. This is especially true of cybersecurity. Analysis of variance is particularly effective in the study of such problems. It is important that in the analysis of variance, each observation is used to simultaneously evaluate all factors and their interaction.

Multi-armed bandit algorithms offer their approach to testing cybersecurity systems, allowing them to more effectively solve the optimization problem and make faster decisions than traditional optimization methods and statistics. In practice, "bandit" algorithms can handle three or more model variants, while making the best choice. As for traditional statistical testing procedures, the efficiency of the selection process in bandit algorithms for three or more variants of control and calculation versions far surpasses them.

With respect to all the methods considered, it is also necessary to pay attention to the control statistics, the metric used to compare Group A with Group B and more. The control group is subject to the same factors (except the target metric) as the test group. Therefore, without a control group, there is no guarantee that "other conditions will be the same" and that any difference in metrics does occur under experimental conditions (or random way).

References

[1] Alpaydin E. Machine learning: a new artificial intelligence / E. Alpaydin. — M.: Tochka Publishing Group, 2017. — 208 p.

[2] Anderson K. Analytical Culture. From data collection to business results / K. Anderson. — M.: Mann, Ivanov and Farber, 2017. — 336 p.

[3] Afifi A. Statistical analysis. An approach using computers / A. Afifi, S. Eisen. — M.: Mir, 1982. — 482 p.

[4] Bidiuk P.I. Applied statistics / P.I. Bidiuk, O.M. Terentyev, T.I. Prosiankin-Zharkov. — Vinnitsa: PP "TD" Edelweiss and K ", 2013. — 288 p.

[5] Brandt Z. Statistical Methods for Analyzing Observations / Z. Brandt. — M.: Mir, 1975. — 313 p.

[6] Bruce P. Applied statistics for Data Science professionals / P. Bruce, E. Bruce. — SPb .: BHV Publishers, 2018 — 304 p.

[7] Volkova V.N. Fundamentals of systems theory and systems analysis / V.N. Volkova, A.A. Denisov. — M.: Higher School, 2006. — 511 p.

[8] Diogenes Y. Cybersecurity: strategies of attack and defense / Y. Diogenes, E. Ozkay. — M.: DMK Press, 2020. — 326 p.

[9] Kachinsky A.B. Security, Threats and Risk / A.B. Kachinsky. — K .: IPNB RNBO; NA SB of Ukraine, 2004. — 472 p.

[10] Lapan M. Deep learning with reinforcement. AlphaGo and other technologists / M. Lapan. — M.: Peter Press, 2020. — 496 p.

[11] Pollard. J. Reference on computational methods of statistics / J. Pollard. — M. Finance and statistics, 1982. — 344 p.

[12] Ravichandiran S. Deep learning with Python reinforcement. OpenAI Gym and TensorFlow for pros / S. Ravichandiran. — M.: Peter Press, 2020. — 320 p.

[13] Sutton R.C. Training with reinforcement / R.S. Sutton., E.G. Barto. — M.: BINOM "Laboratory of Knowledge". BHV, 2018 — 304 p.

[14] Sulitsky V.N. Methods of statistical analysis in management / V.N. Sulitsky. — M.: Delo, 202. — 520 p.

[15] Chio K. Machine learning and safety / K. Chio, D. Freeman. — M.: DMK Press, 2019. — 388 p.

[16] Schaeff G. Analysis of variance / G. Schaeff. — M.: Science, 1980. — 512 p.