Stochastic Dichotomous Game-Theoretic Model of Technology Efficiency

Vladimir Tsyganov

Institute of Control Sciences, Profsoyuznaya str. 65, 117997 Moscow, Russia
bbc@ipu.ru

Abstract. Theoretical, algorithmic and methodological aspects of stochastic modeling and technology efficiency control based on game-theoretic approach and machine learning are considered. The problem of assigning one of two ranks to a control object with the stochastic potential of technology is posed and solved for the case when probabilistic characteristics are known. Otherwise, to determine the optimal parameter of the ranking rule it is proposed to use the procedure of machine learning. The case of asymmetric awareness of the manager deciding on the ranking and the staff responsible for the effectiveness of the technology is considered. Far-sighted staff selects indicators of technology efficiency in such a way as to maximize own objective function, which depends on current and future results of ranking. There is a game between staff and manager which can lead to a decrease in the effectiveness of the technology and distortion of the estimates of the ranking parameters. This makes machine learning ineffective. To solve these problems, stochastic game model and ranking learning mechanism are proposed. The results of this mechanism functioning are estimates of ranking parameters, standards and ranks that determine staff stimuli. Sufficient conditions for the synthesis of ranking learning mechanism have been found, allowing to reveal the potential of technology effectiveness and to determine the optimal parameters of the ranking rule. These conditions are illustrated by the example of machine learning of ranking the technology electricity effectiveness in the process of implementing the program to increase the energy efficiency of the Russian Railways holding.

Keywords: stochastic modeling, game theory, control, machine learning, ranking, stimulation

1 Introduction

The important aspect of production efficiency in the face of changes is the disclosure of internal reserves and technological resources by activating the human factor. This determines the relevance of the game-theoretic approach to stochastic modeling of the dynamics of organizational and technical systems. This approach allows take into account conflicts of interest and the resulting activity of people in the production process.

The problem of reconciling these interests under conditions of uncertainty has traditionally been considered in hierarchical games, or inverse Stackelberg games first introduced in [1]. In Russian/Soviet literature, hierarchical games were previously considered by Yu. Germeier [2].

At the heart of modern research of organizational systems control, taking into account the activity of their elements is a game-theoretic approach. Among the scientific directions, it should be noted first of all, the theory of mechanisms (mechanism design), which has incorporated the control theory, the theory of contracts and the theory of feasibility [3]. It should also be noted theory of active systems [4], as well as works on the study of individual models using the game-theoretic approach for example [5].

On the other hand, at present, the theoretical, algorithmic and methodological aspects of stochastic game-theoretic modeling of organizational and technical systems with elements of artificial intelligence are acquiring more importance. For example, within the framework of the theory of active systems, mathematical methods are developed for smart organization mechanism design [4].

The most important element of artificial intelligence is learning. Accordingly, game-theoretic approach support mechanism design with learning. For example, a behavioral model for mechanism design is based on individual evolutionary learning [6].

In recent years, machine learning has attracted a lot of interest from both scientists and managers of organizational and technical systems. But the development of machine learning at the beginning of the 21st century has led to a certain divergence with control theory [7]. Control-related views have only been published for narrower areas of iterative learning by Bristow et al. [8] and of reinforcing learning by Recht [9]. What is a challenge for control theorists is that there is very little rigorous mathematical proof in the new machine learning "tsunamis" [7]. Today this divergence is the subject of discussion in the control community on "Control and learning – is there really a divide?" Among the issues currently being discussed is: How can we use control theory to improve machine learning algorithms? [10].

To answer these questions in relation to control of organizational and technical systems, mechanisms have been designed in recent years using machine learning algorithms. So, in [11] on the basis of a game-theoretic approach the mechanism for learning digital control of a large-scale industrial system was designed. In [12], a business control mechanism based on a learning algorithm with a tutor proposed. In [13], a control mechanism of production was designed based on a self-learning of identification. This paper discusses a game-theoretic approach to designing a mechanism that uses a self-learning algorithm for ranking and stimulation in organizational and technical system.

2 Stochastic Dichotomy

Many tasks of governing body (briefly – Center) come down to assigning one of two ranks to the control object (briefly – dichotomous ranking). The better the rank, the higher the incentive for the controlled object. For this, a decision rule is needed. With enough complete a priori information, Center can use the rules of the theory of statistical decisions.

2.1 Minimization of Losses in Dichotomy

Denote by t the time period, $t = 0,1,...$ The sequence $t = 0,1,...$ is a chronologically sorted sequence of time intervals. These intervals are constant in length, not overlapping and cover the whole timeline. Unit of time is used here correspond to the application problem of organization control is considered (for example month, see Section 4).

Let z_t be a random variable characterizing the efficiency potential of the object in period t, $z_t \in \Delta$, where $\Delta = [\lambda, \sigma]$ is the finite subset of $R^1 : \Delta \subset R^1$, $\lambda < 0, \sigma > 0$. Variables z_t corresponding different periods $t = 0,1,...$, are independent identically distributed random variables. Dichotomous ranking involves z_t assigning one of the two areas that make up the set Δ. Incorrect ranking leads to losses.

Denote $\{\Delta_1, \Delta_2\}$ some partition of the set Δ into 2 areas, $\bigcup_{k=1}^{2} \Delta_k = \Delta$. When ranking, i.e. assigning a situation z_t to one of these areas, Center makes a decision associated with some losses. The challenge is to define a partition $\{\Delta_1, \Delta_2\}$ that minimizes the average losses associated with ranking. We introduce for each, so far unknown area Δ_k , $k = \overline{1,2}$, the dichotomous ranking loss function:

– $L_1(c,z)$ – losses in case of assignment z the rank 1, while $z \in \Delta_2$;

– $L_2(c,z)$ – losses in case of assignment z the rank 2, whereas $z \in \Delta_1$,

where c is an unknown parameter. Then the affiliation z of one or another area is determined by the sign of the decision rule

$$\mu_{12}(c,z) = L_1(c,z) - L_2(c,z): \; z \in \Delta_1 \text{ for } \mu_{12}(c,z) < 0, \text{ and } z \in \Delta_2 \text{ for } \mu_{12}(c,z) \geq 0. \quad (1)$$

Assume that Center knows the distribution density $q(z)$ of a random variable z. Then the problem is solved by determining the parameter c of the decision rule (1) that minimizes the average losses of ranking:

$$J(c) = \sum_{k=1}^{2} \int_{\Delta_k} L_k(c,z) q(z) dz \mapsto \min_c \quad (2)$$

2.2 Learning of Stochastic Dichotomy

Knowing $q(z)$ we can find the parameter as a solution to problem (2). However, in a stochastic setting a priori information is often not enough. Suppose $q(z)$ it is unknown to Center. Therefore, the direct determination of the parameter c by solving optimization problem (2) is impossible. Then we can then try to tune decision rule parameter c using observations z_t, $t = 0,1,...$, to minimize losses $J(c)$.

Write the condition for the minimum average losses of ranking (2) in the form:

$$M_z\left\{ \sum_{k=1}^{2} \hat{B}_k(c,z)\partial L_k(c,z)/\partial c \right\} = 0\,, \tag{3}$$

$$\hat{B}_k(c,z) = \begin{cases} 1 & \text{if} \quad z \in \Delta_k \\ 0 & \text{otherwise} \end{cases}, \tag{4}$$

where M_z is the mathematical expectation operator. For simplicity of calculations, consider below the linear loss functions:

$$L_1(c,z) = z - vc, \quad L_2(c,z) = d(c-z), \tag{5}$$

where

– v is the parameter of loss elasticity when z is assigned rank 1, while $z \in \Delta_2$, $0<v<1$;

– d is the parameter of loss elasticity when assigning rank 2 to z, whereas $z \in \Delta_1$, $d>0$.

Substituting (5) in (1), we obtain the decision rule in the form:

$$z \in \Delta_1 \ \text{ if } \ z < (d+\nu)c/(d+1), \ \text{ and } \ z \in \Delta_2 \ \text{ if } \ z \geq (d+\nu)c/(d+1) \tag{6}$$

Solve equation (3) using the method of stochastic approximation [14]. Denote c_t the optimal estimate of unknown parameter c in period t, obtained by this method. Then, according to (3) we obtain a recursive equation for such estimate:

$$c_{t+1} = c_t - \gamma_t \sum_{k=1}^{2} \hat{B}_k(c_t,z_t)\partial L_k(c_t,z_t)/\partial c_t, \ \ c_0 = b_0, \ \ z_0 = g_0, \ \ t = 0,1,... \tag{7}$$

There b_0 is the initial value of unknown parameter estimate, g_0 is the initial value of the efficiency potential, γ_t is the adaptation coefficient in period t, $\gamma_t > 0$, $\sum_{t=0}^{\infty}\gamma_t < \infty$ [14]. Substituting (4), (5) into (7), and taking into account (6), we obtain the algorithm for optimal estimate the parameter of the decision rule

$$c_{t+1} = I_t(c_t,z_t) = \begin{cases} c_t + \gamma_t v & \text{if} \quad z_t < (d+v)c_t/(d+1) \\ c_t - \gamma_t d & \text{if} \quad z_t \geq (d+v)c_t/(d+1) \end{cases}, \ \ c_0 = b_0, \ \ z_0 = g_0, \ \ t = 0,1,... \tag{8}$$

$$c_{t+1} = I_t(c_t,z_t) \underset{t}{\rightarrow} c^* = \underset{c}{\operatorname{argmin}}\, J(c) \tag{9}$$

3 Game-Theoretic Approach to Machine Learning of Dichotomy

The elements of production are people, as well as the technologies and control processes by which they carry out their activities. Therefore, when researching and developing organizational and technological systems in the face of uncertainty, it is necessary to take into account the socio-psychological aspects of production activities including undesirable activity of staff.

3.1 Stochastic Technological Active System

Let us consider a two-level stochastic technological active system, at the upper level of which Center is located, and at the lower level is forward-thinking staff (FTS) that implements the technological process. Let us characterize game-theoretic approach to the study of the interaction of Center and FTS to increase efficiency of the technology, using the results of Section 2.

Suppose that random value of technological potential z_t becomes known FTS before choosing indicator y_t of technology efficiency in period *t*. Moreover, value z_t is unknown to Center. Based on the condition that the efficiency indicator cannot exceed the potential (i.e. $y_t \le z_t$), FTS chooses y_t, $t = 1, 2, ...$, in such a way as to maximize own target function, depending on current and future ranks assigned by Center.

Center, on the other hand, observes only indicator y_t that does not necessarily coincide with potential z_t (since $y_t \le z_t$). Therefore, Center is forced to form the rank of FTS under conditions of uncertainty caused not only by stochastic potential of the technology, but also by the undesirable activity of FTS. For this, Center uses the learning algorithm (8), substituting the observed indicator y_t in it, instead of the unknown z_t. In this case, the estimate a_{t+1} of the parameter of the decision rule is obtained using algorithm similar to (8):

$$a_{t+1} = I_t(a_t, y_t) = \begin{cases} a_t + \gamma_t v & \text{if} \quad y_t < (d+v)a_t/(d+1) \\ a_t - \gamma_t d & \text{if} \quad y_t \ge (d+v)a_t/(d+1) \end{cases}, \; a_0 = b_0, \; y_0 = g_0, \; t = 0,1,... \quad (10)$$

3.2 Goals of Center

Center is interested in unlocking the potential of a technology, as well as in increasing its predictability. Constant disclosure of technology potential is achieved when $y_t = z_t$, $t = 0,1,...$ Unpredictability of this potential is due not only to the random factors, but also to undesirable activity of FTS, due to which $y_t < z_t$.

Since in general case $y_t \ne z_t$, then $a_t \ne c_t$, $t = 1, 2, ...$ Therefore, estimate a_t calculated using recurrence algorithm (10) does not converge to optimal estimate $c*$ determined according to (9). This makes machine learning algorithm (10) ineffective. Substantially, the reason is that Center is not able to take into account random factors that become known to FTS in the process of production. This not only reduces the effectiveness ($y_t < z_t$), but also makes such machine learning inefficient. To improve the efficiency of machine learning, it is necessary to ensure convergence of estimate a_t calculated by recurrent algorithm (10) to optimal estimate c^* determined according to (9):

$$a_{t+1} = I_t(a_t, y_t) \underset{t}{\rightarrow} c^* = \arg\min_c J(c) \tag{11}$$

Thus, Center's goals are to unleash potential of technologies ($y_t = z_t$, $t = 1, 2, ...$), as well as to increase efficiency of machine learning by implementing (11). Thus, expected payoff of Center is maximal if $y_t = z_t$, $t = 1, 2, ...$ To achieve this, Center establishes the necessary order and mechanism for functioning of technological active system.

3.3 Dichotomous Ranking Learning Mechanism

Consider the following order of system's functioning in period t, $t = 0, 1...$ In period $t = 0$ Center knowing initial values a_0 and y_0 calculates estimate a_1 for period 1 by means of (10). Then Center reports a_1 to FTS.

In period t, $t = 1, 2, ...,$ FTS knows not only a_t, but also z_t, $z_t \in \Delta$. Based on this, FTS chooses indicator y_t^* that is preferable for itself, $y_t^* \le z_t$. Then Center, based on observation of y_t^* and known estimate a_t, determines rank FTS $r_t = R(a_t, y_t^*)$. Also Center calculates estimate a_{t+1} for next period t+1 by means of (10). Then Center reports a_{t+1} to FTS, $t = 1, 2, ...$

We will call learning procedure a one-way infinite sequence of functions $I_t(a_t, y_t)$ defined according to (10) with $t = 0, 1, ...,$ and denote it

$$I = \left\{ I_t(a_t, y_t),\ t = 0, 1, ... \right\} \tag{12}$$

Similarly, we will call $R = \left\{ R(a_t, y_t),\ t = 0, 1, ... \right\}$ ranking procedure. Then learning procedure I and ranking procedure R are combined into a dichotomous ranking learning mechanism $\Sigma = \{I, R\}$ in two-level technological active system shown in fig.1.

3.4 Target of Forward-Thinking Staff

Knowing a_t, z_t, and Σ, FTS chooses y_t in such a way as to increase own target function V_t, which depends on current and future ranks $r_\tau = R(a_\tau, y_\tau)$, $\tau = \overline{t, t+T}$:

$$V_t(\Sigma) = \sum_{\tau=t}^{t+T} \rho^{\tau-t} R(a_\tau, y_\tau) \tag{13}$$

where ρ is the discount rate, $0 < \rho < 1$, T is the number of periods taken into account by FTS.

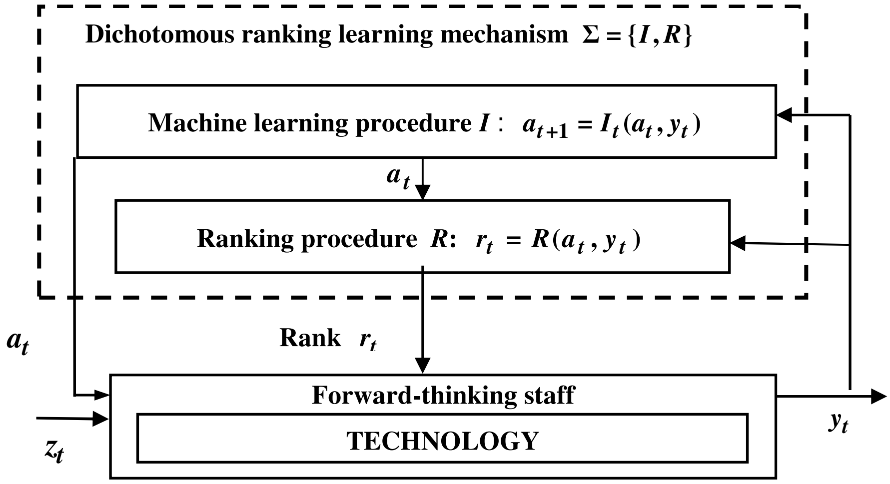

Fig.1. Technological active system with dichotomous ranking learning mechanism Σ .

We will assume that FTS know that $0 \le y_\tau \le z_\tau$, and $z_\tau \in \Delta$ in each future period τ, $\tau = \overline{t+1, t+T}$. In order to make a choice in conditions of such uncertainty, FTS is guided by expected payoff equals to guaranteed value of (13):

$$w_t(a_t, y_t, z_t, \Sigma) = \min_{z_\tau \in \Delta,\ \tau = \overline{t+1, t+T}} \ \max_{0 \le y_\tau \le z_\tau,\ \tau = \overline{t+1, t+T}} V_t(\Sigma), \quad t = 1, 2, ... \qquad (14)$$

Then the set of possible choices of FTS has the form:

$$W_t(a_t, z_t, \Sigma) = \{ y_t^* \mid 0 \le y_t^* \le z_t,\ w_t(a_t, y_t^*, z_t, \Sigma) \ge w_t(a_t, y_t, z_t, \Sigma),\ 0 \le y_t \le z_t \}, t = 1, 2, ... (15)$$

Suppose that the set of FTS possible choices (15) includes z_t, i.e. $z_t \in W_t(a_t, z_t, \Sigma)$. In this case we say that the benevolence hypothesis of FTS with respect to Center is valid if FTS chooses $y_t^* = z_t$, $t = 1, 2, ...$

3.5 Synthesis of Dichotomous Ranking Learning Mechanism

Let us turn to current practice of production management. First, ranking and stimulus procedures are usually designed in such a way stimuli grow as production indicators increase compared with their current scores (plans, standards). Usually, stimulation is carried out in case these scores are exceeded [4]. Therefore, the higher a score the more difficult it is to get a stimulus.

Secondly, forecasting procedure in a large corporation is often organized in such a way that a score (like plan) in each subsequent period increases by a certain percentage of result achieved today (the so-called "planning from the achieved level") [4]. Then future score (plan, standard) will be the higher, the higher today's indicators are achieved. Therefore, staff may not be interested in exceeding the score (after all, the higher the score in the future the more difficult it will be to get a stimulus).

Thus, problem arises of the lack of interest of forward-thinking staff in unlocking the potential of technology. In this case $y_t < z_t$, and it is impossible to determine c^* with the aid of machine learning algorithm (10). Therefore, expected payoff of Center is not maximal if $y_t^* < z_t$, $t = 0,1,...$

Consider the non-cooperative game of Center and FTS in two-level technological active system shown in fig.1. The first move in period t is made by Center, setting mechanism $\Sigma = \{I, R\}$. The second move is made by FTS, choosing indicator y_t^*. Then these moves are repeated in next period $t+1$, $t = 1,2,...$

Statement. To unlock the potential ($y_t^* = z_t$, $t = 1,2,...$) and improve the efficiency of machine learning to obtain (11), it is enough to set a mechanism $\Sigma = \{I, R\}$ with procedure I that satisfies (12), and

$$R(a_t, y_t) = \Theta(y_t - x_t), \quad x_t = a_t(d+v)/(d+1), \tag{16}$$

$$\Theta(y_t - x_t) = \begin{cases} 2 & \text{if } y_t \geq x_t, \\ 1 & \text{if } y_t < x_t \end{cases} \tag{17}$$

Proof. The expected payoff of FTS (14) depends according to (13) on both current and future ranks $r_\tau = R(a_\tau, y_\tau)$, $\tau = \overline{t, t+T}$. By condition (16), with an increase in indicator y_t, current rank FTS $r_t = R(a_t, y_t)$ increases (does not decrease). In addition, by the hypothesis of Statement, Center uses learning procedure (12). Hence, according to (10) estimates a_τ decrease (do not increase) with an increase in y_t, $\tau = \overline{t+1, t+T}$. Therefore, according to (16) future ranks FTS $r_\tau = R(a_\tau, y_\tau)$ increase (do not decrease) with an increase of y_t, $\tau = \overline{t+1, t+T}$.

According to (13), $V_t(\Sigma)$ increases monotonically in $r_\tau = R(a_\tau, y_\tau)$, $\tau = \overline{t, t+T}$. But r_t monotonously increases (does not decrease) by y_t. Therefore, with an increase in indicator y_t, expected payoff (14) also increases (does not decrease). Since $y_t \leq z_t$, maximum of $w_t(a_t, y_t, z_t, \Sigma)$ is reached at $y_t = z_t$. Therefore, according to (15), $z_t \in W_t(a_t, z_t, \Sigma)$. Hence, by virtue of the benevolence hypothesis, $y_t^* = z_t$, $t = 0,1,...$ But then (8) and (10) coincide. Therefore, (11) follows from (9), Q.E.D.

Following the common game-theoretic notation [15], let comment relation between this Statement and Nash equilibrium. In our case, Nash equilibrium is a solution of a non-cooperative game in which both Center and FTS know equilibrium strategies of the other player, and no player has anything to gain by changing only own strategy. Center strategy Σ determines actions based on what it has seen happen so far in the game. Expected payoff of Center is maximal if $y_t^* = z_t$, $t = 1, 2, ...$ FTS makes the strategy choice y_t—its own action based on Σ and z_t. Expected payoff of FTS is (14). According to Statement, no player can increase expected payoff by changing strategy while the other player keep own strategy unchanged. Therefore the set of strategy choices $\{\Sigma, y_t^* \mid t = 1,2,...\}$ constitutes Nash equilibrium. This unleashes technology potential and increases efficiency of machine learning

Consider a simple interpretation of Statement. Suppose that a performance stimulation is such that the higher the rank the higher the stimulus for staff. Center observes a value y_t^* characterizing actual effectiveness in period t, $y_t^* \le z_t$, where z_t is unknown random maximal efficiency. Center learns tuning the decision rule parameter with the aid of (10). Further, in accordance with adopted decision rule, Center ranks FTS according to actual effectiveness of technology. Namely, with $y_t^* \ge x_t$ FTS refers to successful ($\Theta = 2$) and is encouraged. If $y_t^* < x_t$ then FTS refers to dysfunctional ($\Theta = 1$) and is punished.

Any of these decisions is associated with a certain losses for Center. In first case, losses L_1 increase with a decrease in efficiency y_t (undeserved encouragement or bonus of staff). In second case, these losses L_2 increase with increasing efficiency and unfair punishment of staff. The standard $x_t = (d + v)a_t / (d + 1)$ corresponds to the lower limit of satisfactory work of staff.

Note that according to (10), the higher is technology efficiency indicator (y_t^*) the lower is estimate for the next period (a_{t+1}). But, according to (17), this estimate plays the role of the threshold value of indicator y_{t+1} at which FTS receives a stimulus in period t+1. Therefore, FTS becomes easier to get a stimulus in period t+1 even with a smaller value of random potential z_t.

In other words, with an increase in indicator y_t^*, staff receives not only a higher stimulus. With growth y_t^*, estimate for the next period a_{t+1} decreases. Therefore threshold value for stimulation in the future decreases. This further interests the staff in unlocking the potential of the technology, i.e. in choosing $y_t^* = z_t$. Thus, in accordance with (8) – (10) learning procedure (12) ensures convergence of estimate a_t to optimal value $c*$ (11). This makes machine learning algorithm (10) more efficient.

4 Example: Using 2 dichotomous ranking learning mechanisms for ranking electricity efficiency in four ranks

By Statement, dichotomous ranking learning mechanism promotes revelation of potential effectiveness of technology and increase efficiency of machine learning. Consider the application of this mechanism to ranking the efficiency of energy-saving technology under the program of increasing the energy efficiency of the Russian Railways holding.

4.1 Ranking Learning Mechanism

Improving energy efficiency of production technology can be achieved through modernization and optimization of technological processes, as well as operating modes of heating and lighting systems of an enterprise. The decision on ranking the efficiency of energy-saving technology is made by the employee of the energy commission of the Russian Railways, responsible for achieving indicators of the program for improving energy efficiency in region. This employee acts as Center, observing actual effectiveness of energy-saving technology at a subordinate enterprise providing wagon-repairing [16]. Consider the application of a mechanism $\Sigma = \{I, R\}$ that satisfies conditions of Statement to improve efficiency of electricity use at such enterprise (briefly, FTS).

The monthly electricity efficiency of wagon-repairing α_t is calculated as number n_t of wagons repaired in month *t*, divided by appropriate number e_t of megawatt-month of consumed electricity: $\alpha_t = n_t / e_t$. There t is the number of month in a year, $t = \overline{1,12}$. Indicator y_t of monthly electricity efficiency is calculated as deviation of α_t from the norm (plan) of electricity efficiency β_t divided by this norm: $y_t = (\alpha_t - \beta_t)/\beta_t$ $t = \overline{1,12}$. Thus, the value of indicator y_t depends on variables n_t, e_t, β_t. In turn, these variables depend on many random factors: volume of orders for repairs, weather, season, etc. Therefore value y_t has a stochastic character.

Based on indicator y_t, a monthly ranking of technology effectiveness is determined. For this, Center uses decision rule with customizable estimates based on algorithm (10) and conditions of Statement. In order to make the results more vivid, Center defines four ranks of electricity efficiency. There rank 4 corresponds to excellent assessment of electricity efficiency, rank 3 – to good assessment, rank 2 – to satisfactory assessment, and rank 1 – to poor assessment.

To determine these four ranks of electricity efficiency, Center uses the following procedure for estimate parameters of decision rule and procedure for ranking, based on the approach developed above.

1. In case indicator y_t is negative ($y_t < 0$) then rank r_t of electricity efficiency in period *t* can be 1 or 2. At the beginning of the year (in period $t = 0$), Center and FTS know d, v, initial values of indicator y_0, estimate a_0, and standard

$x_0 = a_0(d+v)/(d+1)$. After that, within a year the estimate of decision rule parameter is adjusted according to algorithm (10), by formula

$$a_{\tau+1} = \begin{cases} a_\tau + \gamma_\tau v & \text{if } y_\tau < x_\tau \\ a_\tau - \gamma_\tau d & \text{if } y_\tau \geq x_\tau \end{cases}, \quad x_\tau = a_\tau(d+v)/(d+1), \quad \tau = \overline{0,12}, \tag{18}$$

Then according to Statement, the rank of electricity efficiency is

$$r_t = \begin{cases} 2 & \text{if } x_t \leq y_t < 0 \\ 1 & \text{if } y_t < x_t \end{cases}, \quad t = \overline{0,12} \tag{19}$$

2. In case indicator y_t is non-negative ($y_t \geq 0$) then the rank of electricity efficiency may be 3 or 4. In this case, procedures for parameter estimates of decision rule and ranking are based on the approach developed above. Namely, similar to (5) denote: a_t^+ – the estimate of parameter of decision rule for ranking of electricity efficiency to 3 or 4 in period *t*; $y_t - va^+$ – the loss function for erroneous assignment of rank 3 (instead of rank 4), $0<v^+<1$; $d^+(a^+ - y_t)$ – the loss function for erroneous assignment of rank 4, $d^+>0$; $x_t^+ = a_t^+(d^+ + v^+)/(d^+ + 1)$; $t = \overline{0,12}$.

At the beginning of the year (in period $t = 0$) Center and FTS know d^+, v^+, initial values of indicator y_0, estimate a_0^+, and standard $x_0^+ = a_0^+(d^+ + v^+)/(d^+ + 1)$. After that, within a year the estimate of decision rule parameter is adjusted similar to (18) by formula:

$$a_{\tau+1}^+ = \begin{cases} a_\tau^+ + \gamma_\tau v^+ & \text{if } y_\tau < x_\tau^+ \\ a_\tau^+ - \gamma_\tau d^+ & \text{if } y_\tau \geq x_\tau^+ \end{cases}, \quad x_\tau^+ = a_\tau^+(d^+ + v^+)/(d^+ + 1), \quad \tau = \overline{0,12} \tag{20}$$

Then in accordance with Statement, the rank of electricity efficiency is calculated using a formula similar to (19):

$$r_t^+ = \begin{cases} 4 & \text{if } x_t^+ \leq y_t \\ 3 & \text{if } 0 \leq y_t < x_t^+ \end{cases}, \quad t = \overline{0,12} \tag{21}$$

Note that (20) and (21) are similar to (18) and (19), if a_t, x_t, r_t replaced by a_t^+, x_t^+, r_t^+, respectively.

4.2 Parameters of Calculations

Estimates a_t and a_t^+ in period t were calculated, respectively, using formulas (18) and (20). Based on these estimates, values of x_t and x_t^+ were determined according to (18) and (20). Values x_t and x_t^+ make sense, respectively, of the lower and the upper standard of electricity efficiency in period t, $t = \overline{0,12}$.

The values of parameters used in calculations of a_t and a_t^+ were based on hard and soft knowledge. Information theory of identification [14] shows that in optimal algorithm $\gamma_t = 1/(t+1)$, $t = 0,1,...$ The other needed values were assigned by expert – invited specialist or tutor [12]. Then these values were adjusted by the employee himself or a manager of a higher level.

In this expert environment, it was generally accepted that the loss in case of mistaken assignment of a higher rank was lower than the loss in case of erroneous assignment of a lower rank. Based on this, parameters of loss elasticity in case of an erroneous assignment of a higher rank were taken equal to 0.15 (i.e. $d = d^+ = 0{,}15$). In case of an erroneous assignment of a lower rank, parameters of loss elasticity were taken equal to 0.10 (i.e. $v = v^+ = 0{,}10$).

In addition, experts a priori considered the deviation of indicator y_t from zero level by more than 5% noticeable. The result of such a deviation should be an assignment to a different rank, and appropriately stimulated. Therefore, the initial values of standards of electricity efficiency were taken equal $x_0 = -0{,}05$, $x_0^+ = 0{,}05$. Hence, according to (19) if $y_0 < x_0 = -0{,}05$ then FTS attributed to a rank 1. Also according to (21) if $y_0 \ge x_0^+ = 0{,}05$, then FTS attributed to a rank 4.

4.3 Estimations and Standards Calculations

Note if $x_0 = -0{,}05$, $x_0^+ = 0{,}05$, then according to (18) and (20), $a_0 = -4{,}6$, $a_0^+ = 4{,}6$. The results of estimations and standards calculations for above parameters are shown in fig.2. There are graphs of indicator y_t, as well as lower and upper standard x_t and x_t^+ during the year, $t = \overline{0,12}$.

According to (20), if indicator y_t is less than the upper standard x_t^+ then this standard is increased. For example, according to fig.2 the upper standard x_t^+ increases in February and March. Also from September until the end of the year, the standard x_t^+ was raised because y_t was less than the standard (here the growth of the lower standard x_t^+ is less noticeable on fig. 2 due to the smallness of the γ_t). On the other hand, according to (21), the standard x_t^+ is a threshold of excellent rank in the future period τ, $\tau > t$. Therefore, even in November, the FTS did not receive an excellent rank, although indicator y_t exceeded 0.1. Note that before September this would have been enough to get an excellent rank.

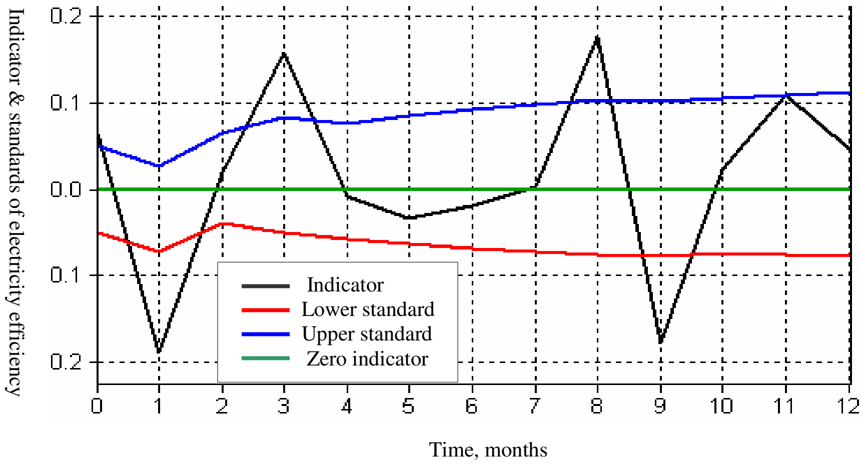

Fig.2. Monthly indicator, lower standard, and upper standard of electricity efficiency.

Also from fig.2 we see growth of lower standard x_t during the second month of the year (in February). The reason is the low FTS indicator in January. This is explained by the fact that according to (18), the smaller is indicator y_t the higher is the lower standard x_τ using as a threshold for satisfactory rank in future period τ, $\tau>t$. In other words, by being poor FTS worsened own ability to become satisfactory rank. The situation is similar in the ninth month (here the growth of the lower standard x_t is also less noticeable on fig. 2 due to the smallness of the γ_t).

However, it was enough for FTS not to get poor rank in the following months (including August) as lower standard x_t began to decline. In essence, from February to August FTS worked for own authority.

4.4 Ranks and Stimuli

Further, Center determined rank R_t in period t, $t=\overline{0,12}$. If current indicator of electricity efficiency was negative (y_t<0) then to assign rank 1 or 2, (19) was used. Otherwise (at y_t≥0) to assign rank 3 or 4, (21) was used. Thus, the procedure for ranking electricity efficiency in period t, combining (19) and (21), had the form:

$$\hat{r}_t = \begin{cases} 4 \quad if \quad y_t \geq x_t^+ \\ 3 \quad if \quad 0 \leq y_t < x_t^+ \\ 2 \quad if \quad x_t \leq y_t < 0 \\ 1 \quad if \quad y_t < x_t \end{cases}, \quad t=\overline{0,12}, \tag{22}$$

where $\hat{r}_t$ is the rank of electricity efficiency in month *t*. In fig.3 shows a graph of rank $\hat{r}_t$ calculated by (22), $t = \overline{0,12}$.

Substantially, standard x_t^+ is the lower limit of electricity efficiency y_t, corresponding to excellent work of staff. Standard x_t is the lower limit of electricity efficiency y_t corresponding to satisfactory performance. When electricity efficiency was below x_t in February and October, the intervention of Center was required.

Let comment on the relation between fig. 2 and fig. 3. Note that in general terms dynamics of $\hat{r}_t$ resembles dynamics of y_t. However fig. 3 gives rougher assessments. This was necessary for a qualitative analysis (for example in a report to the higher management when there was no time to go into details). Whereas fig. 2 provides more accurate quantitative estimates of the ratio of indicators and standards useful for in-depth analysis. For example fig. 3 shows that in November (t =11) the rank was 3. However, fig. 2 shows that this month indicator y_t was very close to upper standard, that is, could get rank 4. This shows more progress in increasing electricity efficiency at the end of the year.

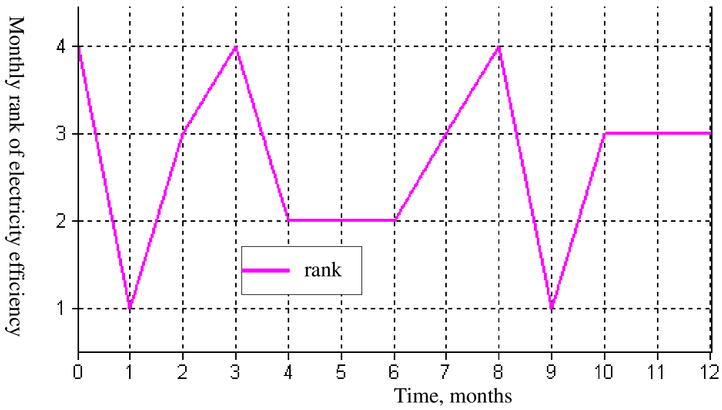

Fig.3. Monthly electricity efficiency ranks of electricity efficiency.

The higher the rank, the more stimuli the staff gets. Denote s_t stimulus for electricity efficiency in period *t*. Formally, the stimulation procedure is such that the stimulus $s_t = S(\hat{r}_t)$ grows monotonously with increasing rank $\hat{r}_t$: $S(\hat{r}_t) \uparrow \hat{r}_t$. Then given (22), the higher the electricity efficiency indicator the higher rank and stimulus of the staff.

In addition, according to (18) and (20) the higher indicator y_t the lower standards for the next period x_{t+1} and x_{t+1}^+. Thus, with an increase in indicator y_t the staff receives not only higher stimulus s_t. Also threshold value for future stimulation x_{t+1}

and x_{t+1}^{+} will decrease. This further piques the interest of staff in maximum disclosure of electricity efficiency potential and makes machine learning more efficient.

5 Conclusions

An important aspect of development and application of game theory and stochastic modeling in a wide range of organizational and technical systems is study of possibility of use of internal reserves and resources of the technologies used. For this, it is necessary to take into account human factor, interests of elements of the system. Often management does not know random interference and other factors that become known to staff in the process of production. This reduces effectiveness of the technology and makes the machine learning inefficient.

The task of the theory was the synthesis of game-theoretic mechanisms of coordinated control in stochastic conditions, in which desire of elements to achieve own interests leads to an increase in technology effectiveness. The method for solving this problem by synthesizing the optimal learning mechanism for dichotomous ranking was proposed. This method includes construction of procedures for stochastic approximation of decision rule, ranking and stimulation. Thanks to this mechanism, Nash equilibrium arises which makes machine learning more efficient.

This approach was illustrated by the example of machine learning to rank electrical efficiency of wagon-repairing aimed at achieving targets established by the program of increasing the energy efficiency of the Russian Railways. Further research will focus on the development of theoretical, algorithmic and methodological aspects of game theory and stochastic modeling with a focus on their application in a wide range of active organizational and technical systems.

References

1. Ho, Y.-C., Luh, P., Muralidharan, R.: Information structure, Stackelberg games, and incentive controllability. IEEE Trans. Automat. Control 26(2), 454–460 (1981).
2. Germeier, Yu.: Igry s neprotivopolozhnymi interesami (Games with non-opposing interests), Nauka, Moscow (1976).
3. Jackson, M.: Mechanism theory. In: Derigs U. (ed.) Optimization and Operations Research, Encyclopedia of Life Support Systems. EOLSS Publishers, Oxford (2003), https://web.stanford.edu/~jacksonm/mechtheo.pdf, last accessed 2020/08/06.
4. Burkov, V., Gubko, M., Kondratiev, V., Korgin, N., Novikov, D.: Mechanism design and management. Mathematical methods for smart organizations. NOVA Publishers, New York (2013).
5. Van Essen, M.: A note on the stability of Chen's Lindahl mechanism. Social Choice and Welfare 38(2), 365-370 (2012).
6. Arifovic, J., Ledyard, J.: A behavioral model for mechanism design: individual evolutionary learning. Journal of Economic Behavior and Organization 78, 375–395 (2011).

7. Fradkov, A.: Early history of machine learning. In: 21st IFAC World Congress, 3439. Berlin (2020).
8. Bristow, D., Tharayil, M., Alleyne, A.: A survey of iterative learning control. IEEE Control Systems Magazine 26(3), 96-114 (2006).
9. Recht, B.: A tour of reinforcement learning: the view from continuous control. In: arXiv:1806.09460v2 [math.OC] 10 November 2018, https://arxiv.org/pdf/1806.09460.pdf, last accessed 2020/08/06.
10. Recht, B.: Reflections on the learning-to-control renaissance. In: 21st IFAC World Congress, 4707. Berlin (2020).
11. Tsyganov, V.: Learning mechanisms in digital control of large-scale industrial systems. In: Global Smart Industry Conference, pp. 1–5. IEEE, Chelyabinsk, Russia (2018).
12. Tsyganov, V.: Tutoring mechanisms of business management. In: 21st International Conference on Business Informatics, vol. 2, pp. 60-67. IEEE, Moscow (2019).
13. Tsyganov, V.: Designing adaptive information models for production management. Procedia CIRP 84, 1088-1093 (2019).
14. Tsypkin, Ya.: Fundamentals of the information theory of identification. Nauka, Moscow (1984).
15. Osborne, M., Rubinstein, A.: Course in game theory. MIT, Cambridge MA (1994).
16. Tsyganov, V.: Decision making and learning in wagon-repairing. In: 12th Conference on Management of Large-Scale System Development. pp. 1–5. IEEE, Moscow (2019).