ISSN 0005-1179 (print), ISSN 1608-3032 (online), Automation and Remote Control, 2023, Vol. 84, No. 4, pp. 474–485.
Trapeznikov Institute of Control Sciences, Russian Academy of Sciences, 2023.

Identification of Critical States of Technological Processes based on Predictive Analytics Methods

S. M. Kovalev[*,**,a], **A. V. Sukhanov**[*,**,b], **I. A. Olgeyzer**[*,**,c], **and K. I. Kornienko**[*,***,d]

[*] *Rostov Branch of JSC "Scientific Research and Design Institute of Informatization, Automation and Communication in Railway Transport," Rostov, Russia*
[**] *Rostov State Transport University, Rostov, Russia*
[***] *Emperor Alexander I St. Petersburg State Transport University, St. Petersburg, Russia*
e-mail: [a]*ksm@rfniias.ru*, [b]*a.suhanov@rfniias.ru*, [c]*iohan@rfniias.ru*, [d]*k.kornienko@vniias.ru*

Received April 27, 2022
Revised August 8, 2022
Accepted November 30, 2022

Abstract—The paper proposes a predictive approach to assessing special classes of dangerous states in the development of technological processes in order to make proactive decisions. The developed approach is based on a hybrid model based on the combination of an evidence-based classifier, fuzzy logic, and a Dempster–Shafer probabilistic scheme for evidence combination. The article presents a formal description of the predictor of critical states of the technological process. The resulting approach is universal and applicable in the automation of any complex technical systems. As an example, this article considers the application of the developed approach to solve the problem for assessing the safety of the technological process of shunting trains on a hump yard. The presented example shows the high efficiency and practical usefulness of the developed approach.

Keywords: complex technological situation, predictive analytics, hump yards, fuzzy logic, and evidence classifier

DOI: 10.25728/arcRAS.2023.28.68.001

1. INTRODUCTION

Assessment of the state of complex technical systems (CTS) has not lost its relevance for many years [1, 2]. The increase in the degree of complexity of CTS requires the development of known and the development of new methods for ensuring the reliability of CTS [3]. Since complex systems are characterized by nonlinearity, a large number of elements, stochasticity, weak structured formalization, etc. [4], their work cannot be modeled with rigid algorithms that are widely used to identify patterns in the implementations of a random process that accompanies the analyzed system, accessible to observation, in order to monitor and control [1]. The most difficult task is to predict target (of interest in a particular area) situations, rather than changes in controlled parameters. The target can be situations of achieving some effect, profit, or the reverse of them—dangerous (abnormal, critical) situations.

This article considers the problem of assessing the occurrence of target critical situations, including CTS failures and crashes. The solution of this problem is a top priority when fulfilling the safety requirements that are increasing every year for the automation of CTS [5]. Timely detection of critical situations creates the possibility of maintaining the working state of the CTS and eliminating the causes of their occurrence.

To solve this problem, a new approach is proposed to identify critical states of technological processes (TP) that accompany the behavior of CTS. The approach is based on a hybrid model that combines a probabilistic model of an evidence-based classifier and a logical-linguistic model in the form of fuzzy production rules.

In the following chapters, a formal statement of the problem of identifying critical states of technological processes is described, a description of the developed approach is presented, and a specific example of the application of the proposed solutions to ensure the safety of the movement of cars at a hump yard is considered.

2. PROBLEM FORMULATION

The main task of the predictive analytics of the TP is to obtain knowledge about the states of the TP and its behavior in the form of descriptions consistent with the knowledge of specialist experts. In this article, we will focus on the development of a new class of predictive models designed for the analysis of TP. These models can be used to predict critical states or emergency technological situations that arise in the process of TP development.

An emergency technological situation (ETS) is considered as a technological situation (TS) associated with the appearance of failures or significant deviations in the development of TP. The ETS requires corrective decisions to be made to normalize the technological process. The developed predictive approach to the TP analysis is based on the idea of detecting special types of ETS precursor events in a controlled process, represented by the corresponding descriptions in the TP model.

The multivariate time series (MTS) act as a TP model:

$$S = (\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_T)$$

TP model characterizes the change in the key parameters of the TP, represented by a set of vector-numerical values $\mathbf{x}_t \in \mathbb{R}^L$ $(t = 1, \ldots, T)$ at time line $[1, T]$. Time line $\{x_{it} \in R | t = 1, \ldots, T\}$ for ith parameter in MTS forms the ith dimension or the component i of the MTS.

ETS precursor events form a special class of predictor events in the development of TP.

At a meaningful, informal level, the task is to develop an approach to detecting predictor pattern events in a time series of numerical data describing the behavior of a technological process. These event-patterns are precursors of emergency technological situations in a controlled process.

Let us clarify the conditions of the problem. First of all, we note that there are a fairly large number of empirically confirmed facts of the existence of predictor events in the areas of environmental monitoring, adaptive control, diagnosing the states of technological and production processes, etc. [1, 6–8]. We analyzed the features of the behavior of the MTS on the eve of the ETS. The description of predictor events and their connection with the ETS is of a weakly formalized nature and can be presented by experts in the form of linguistic descriptions [7]. Particularly, on the eve of bifurcations in nonlinear dynamical systems, the emergence of a special nature of parametric oscillations of the evolving model is observed. This is a key sign of the transition of the model to a new one, including the critical state, which includes the ETS [8]. In the field of rail transportation, specialists also revealed the existence of a causal relationship between ETS and predictor events. Experts in the form of fuzzy-linguistic descriptions can also present predictor events [7]. Another feature of the identification of predictor events is the need to use evolving adaptive models to generate a modeling MTS. Adaptation of the model in the course of TP development makes it possible to reveal the features of its behavior in the run-up to the ETS.

Thus, it is possible to formulate a number of restrictions imposed on the developed predictor model. Firstly, the fuzzy description of predictor events and the fuzzy-defined nature of their connection with the ETS necessitates the use of a class of fuzzy-logical models. This class of models is able to operate with fuzzy and linguistic descriptions when processing the MTS. Secondly, the fuzzy model must be adaptable to the input data in order to generate the dynamics of the simulated TP and identify predictor patterns. Thirdly, the predictor model must include a stochastic component to obtain probabilistic estimates for the detection of predictor events. Based on the foregoing, the problem to be solved is reduced to:

Development of a hybrid logical-linguistic model, including a fuzzy-logical and probabilistic component. This model must be capable identify predictor events in $\boldsymbol{S}$ *with a known degree of probability for a given class of TP, presented in the form of MTS* $\boldsymbol{S}$*, and class ETS, presented in the form of linguistic descriptions or target values of features* $\mathbf{x}_t \in \mathbb{R}^L$.

In the above formulation, a number of points need further clarification, which are given in the subsequent sections.

3. GENERAL APPROACH TO PREDICTIVE PROCESS ANALYSIS

The developed predictive approach to the TP analysis is based on the idea of detecting special types of predictor events in a controlled process. These events precede the appearance of the ETS and should be represented by appropriate descriptions in the TP model. A MTS is used as a TP model, characterizing the change in key TP parameters over time.

We proceeded from the basic assumption that there is a direct relationship between the probabilities of the occurrence of ETS and estimates of the complexity of technological situations that cause these states, when developing a predictive model. This assumption is based on the opinion of technology experts that in a more complex situation it is more difficult to choose an effective solution. Therefore, the probability of making inefficient or even erroneous decisions increases. These decisions subsequently lead to significant violations in the development of the technological process and ETS.

The dependence of the critical states of the TP on the complexity of the TS allows us to approach the assessment and prediction of the states of the controlled process through the estimation of the complexity of the corresponding TS. To implement this approach, it is necessary to have a way to assess the complexity of the TS based on the analysis of the features of their representation in the MTS.

Since the concept of “complexity,” as one of the properties of the analyzed object, is weakly formalized, for its presentation, we use a hybrid model that combines an evidence-based classifier and systems of fuzzy rules. The evidentiary classifier is designed to obtain probabilistic estimates of state hypotheses. Systems of fuzzy rules are designed to evaluate the parameters of probabilistic equations in evidentiary classifiers based on the analysis of the linguistic values of features in the description of the TS. Such a model can be called a classification logical-linguistic model (LLM).

The basis of the LLM is an evidence-based classifier [9], based on the combination of the log regression model [10] and the general Dempster–Shafer evidence combination scheme [11].

4. MODEL OF AN EVIDENTIARY CLASSIFIER

The evidence-based classifier is based on a classification model based on logistic regression (log model). The log model is designed to solve the problems of classifying I-dimensional feature vectors $\mathbf{x} = (x_1, \ldots, x_I)$ by K classes $Q = \{q_1, \ldots, q_K\}$ based on the analysis of linear regression equations (log regression equations) $w_k = \alpha_k \mathbf{x} + \beta_k$ associated with the classes $q_k \in Q$ and $\mathbf{x}$ variables. For the case of two classes $Q = \{q_1, q_2\}$, there is a binary model of logistic regression. The probabilities of class membership hypotheses based on the log model are determined in terms of affine functions of $\mathbf{x}$ using the expression:

$$P_k(\mathbf{x}) = \frac{\exp\left(\boldsymbol{\alpha}_k^T \mathrm{x} + \beta_\mathrm{k}\right)}{\sum\limits_{l=1}^{K} \exp\left(\boldsymbol{\alpha}_l^T + \beta_l\right)}.$$

The log-model classifier is an adaptive model whose parameters are determined based on training using experimental data. For a given training set of examples $\{(\mathbf{x}_t q_t)\}_{t=1}^n$ the parameters $\boldsymbol{\alpha}_k$ and β_k are estimated by maximizing the conditional log-likelihood criterion:

$$\sum_{t=1}^{n}\sum_{k=1}^{K}\left[\delta^k_{q(\mathbf{x}_i)}\ln P_k\left(\mathbf{x}_t\right)+\left(1-\delta^k_{q(\mathbf{x}_t)}\right)\ln\left(1-P_k\left(\mathbf{x}_t\right)\right)\right], \tag{1}$$

where $q(\mathbf{x}_t)$—membership class number for $\mathbf{x}_t$; $\delta^k_{q(\mathbf{x}_t)}$—Kronecker symbol.

The evidence-based classifier model is based on combining the classical log model discussed above and the Dempster–Shafer evidence combination methodology (DS-rules/schemes). In this model, it is assumed that numeric features $x_i \in \mathbf{x}$ included in vector $\mathbf{x}$ act as independent evidence in favor of one or another hypothesis of class membership $q_k \in Q$. Therefore, in contrast to the classical log model, in the evidentiary classifier, the linear regression equation $w_{ik} = \beta_{ik}x_i + \beta_{0k}$ is formed separately for each feature variable $x_i \in \mathbf{x}$ and class $q_k \in Q$, and characterizes the conditional probability $P(q_k|x_i)$ (the quotient probability) that the vector $\mathbf{x}$ belongs to the class q_k for a given value of the attribute x_i. Partial probabilities are calculated for each variable x_i and class q_k as follows:

$$P\left(q_k|x_i\right)=\frac{1}{1+\exp-(\alpha_{ki}x_i+\beta_k)}. \tag{2}$$

The probabilities calculated on the basis of (1) are interpreted in the demonstrative classifier as probabilistic masses $m^k_i(x_i)$ of the class membership hypotheses. These hypotheses are combined based on the DS-rule R_{DS} of combining evidence $R_{DS}: \bigcup_{i=1}^{I} m^i_k(x_i) \longrightarrow m^{\cup}_k(\mathbf{x})$. As a result, the final (unconditional) probabilistic masses $m^k_{\cup}(\mathbf{x})$ of hypotheses of class membership are obtained, interpreted as probabilities of classes $P(q_k|\mathbf{x})$. The demonstrative classifier model is described in detail in [10].

The disadvantage of the classical model of an evidentiary classifier is its limited application for recognizing classes in the space of interdependent features and the inability to describe complex, artsy areas in the feature space.

For example, in one of the car retarder control models, the relationship between the complexity of the TS and the speed of the car is described by linear regression. The complexity of the TS increases with an increase in the speed of the car. But the nature of this dependence is also affected by the speed of the next car. At the same time, with an increase in the speed difference, the complexity of the TS will increase. For example, when two cars move at the same speed, the complexity of the TS will be the same. If the speed of only the first car decreases, the complexity of the TS will increase, since there is a possibility that the first car will overtake the second, which is a dangerous situation.

To remove these limitations, a hybrid model of a logical-linguistic classifier (LLM) is considered below, which is free from these disadvantages.

5. CLASSIFICATION LOGICAL-LINGUISTIC MODEL

The main component of the hybrid LLM is a new type of fuzzy Takagi-Sugeno model (TS-model), which includes production rules of the form:

$$R^j_{ik}:\ IF\ \left(x_1=\mu^j_1\right)\ AND\ldots AND\ IF\ \left(x_I=\mu^j_I\right)\ THEN\ w^j_{ik}=\alpha^j_{ik}x_i+\beta^j_{ik},$$

where $j \in [1, J]$ is the rule number; $i \in [1, I]$ is the variable number; $k \in [1, K]$ is the class number; μ^j_i is the membership function (MF) of the fuzzy term for the ith variable in the jth rule; α^j_{ik} and β^j_{ik} is the log regression equation parameters.

The formalized representation of a fuzzy rule is the following expression:

$$R_{ik}^{j}: \bigwedge_{i=1}^{I} \mu_i^j(x_i) \ \Rightarrow \ w_{ik}^j = \alpha_{ik}^j x_i + \beta_{ik}^j. \tag{3}$$

The degree of association of input vector $\mathbf{x}$ with a fuzzy rule is determined through T-norm [12], which is usually represented by the following production operator:

$$\tau^j(\mathbf{x}) = T_{p=1}^{I} \mu_{pk}^j(x_p) = \mu_{1k}^j(x_1) \times \cdots \times \mu_{Ik}^j(x_I). \tag{4}$$

In the context of the problem being solved, the left and right parts of the TS-model fuzzy rules have the following interpretation. Antecedents of fuzzy rules $A_j = \bigwedge_{i=1}^{I} \mu_i^j$ are linguistic descriptions of technological situations. The consequents of fuzzy rules are linear dependencies w_{ik}^j, which have the meaning of log-regression equations w_{ik}^j, establishing a relationship between the probabilistic estimates of class membership $P(q_k|x_i)$ and the values of parameter $x_i \in \mathbf{x}$.

Thus, in the LLM knowledge base, each jth fuzzy rule is related to some jth scenario of a technological situation and characterizes the relationship between its class membership (probability of belonging to one or another class) and the change in the ith parameter of a given technological situation. At the same time, the description of the technological situation is presented in the antecedent of fuzzy rule R_j in the form of a conjunction of the MF of fuzzy terms. Fuzzy rule R_j is activated by substituting specific values of parameters $x_p^* \in \mathbf{x}^*$ $(p = 1, \ldots, I)$, characterizing the jth TS into antecedent A_j. As a result, degree of antecedent truth $\tau^j(\mathbf{x}) = T_{p=1}^{I} \mu_{pk}^j\left(x_p^*\right)$, is calculated, which in the framework of LLM has the meaning of the degree of correspondence of vector $\mathbf{x}^*$ to the jth situation. The LLM knowledge base forms set $\{R_j\}$ of fuzzy rules, so each input vector $\mathbf{x}^*$ activates not one, but simultaneously a group of fuzzy rules $\{R_j'\} \subseteq \{R_j\}$ with the same antecedents. By combining these rules, based on Takagi-Sugeno's fuzzy inference, a single log-regression dependence $w_{ik}^{\cup}$ is formed, which generalizes the properties of all particular log-regressions $w_{ik}^{j'}$, which are antecedents of the fuzzy rules of group $\{R_j'\}$. Based on the generalized log regression $w_{ik}^{\cup}$, for each specific value of the parameter $x_i^* \in \mathbf{x}$, the conditional probability of class membership $P(q_k|x_i^*)$ is calculated for a given value of the parameter x_i^*.

The fuzzy inference formula for the generalized log-regression dependence $w_{ik}^{\cup}$ based on the subset of fuzzy rules $\{R_j'\}$ has the form:

$$w_{ik}^{\cup} = \frac{\sum\limits_{j}^{|R_j'|} \tau^j(\mathbf{x})\, w_{ik}^j}{\sum\limits_{j}^{|R_j'|} \tau^j(\mathbf{x})}, \tag{5}$$

where $\tau^j(\mathbf{x})$ is the the degree of activation of fuzzy rule $R_j \in \{R_j'\}$, calculated on the basis of (4); $|R_j'|$ is the number of fuzzy rules in group $\{R_j'\}$; w_{ik}^j is the particular log regression in the consequent of fuzzy rule $R_j \in \left\{R_j'\right\}$.

Thus, the fundamental feature of the proposed version of the fuzzy Sugeno model is the inclusion in it of two components that independently support the stochastic and fuzzy-logical concepts. Fuzzy rules antecedents support the fuzzy-logical concept. These antecedents describe at the linguistic level scenarios of technological situations presented by expert technologists. Fuzzy rules consequents support the stochastic concept. The consequents represent log-regression dependencies, based on which the probabilistic estimates of the class hypotheses are calculated.

When the input of the model is the vector of parameters of the technological situation $\mathbf{x} = (x_1, \ldots, x_I)$, based on the fuzzy inference, a generalized log regression w_{ik}^j is formed, which is a kind of synergistic combination of technologically oriented particular log regressions w_{ik}^j. Formulas (3)–(5) are basic for calculating the conditional probabilities of class $P(q_k|x_i)$. It is important that the obtained probability estimates are independent, since they are "generated" by different antecedents of fuzzy rules. This makes it possible to correctly apply the Dempster-Shafer scheme to combine partial probabilities $P(q_k|x_i)$ into single probability estimate $P(q_k|\mathbf{x})$ of the class membership hypothesis.

6. AN EXAMPLE OF DATA PRESENTATION BASED ON LLM

The simplest example of the model described above is the following LLM, designed to represent a "broken" area of one of the two classes q_1, q_2 to be recognized using a single log regression equation $w = \alpha x + \beta$ on numerical axis X:

- IF $x > 0$ AND x is Big THEN $w = \alpha x + \beta$ (Probability q_1 increases with increasing X)
- IF $x > 0$ AND x is Small THEN $w = -(\alpha x + \beta)$ (Probability $q_1$decreases with increasing X)
- IF $x < 0$ AND $|x|$ is Big THEN $w = -(\alpha x + \beta)$ (Probability $q_1$increases with modulus increasing X)
- IF $x < 0$ AND $|x|$ is Small THEN $w = \alpha x + \beta$ (Probability q_1 decreases with modulus increasing X)

The formal representation of LLM is the following system of fuzzy rules:

$$\left(\frac{\operatorname{sgn}(x)+1}{2}\right) \& \mu_{Big}(x) \rightarrow w = ax + b$$

$$\left(\frac{\operatorname{sgn}(x)+1}{2}\right) \& \mu_{Sm}(x) \rightarrow w = -(ax + b)$$

$$\left(\frac{-\operatorname{sgn}(x)+1}{2}\right) \& \mu_{Bigmod}(x) \rightarrow w = (ax + b)$$

$$\left(\frac{-\operatorname{sgn}(x)+1}{2}\right) \& \mu_{SMmod}(x) \rightarrow w = -(ax + b)$$

The formal representation for linguistic terms are determined through the parameters of the log-regression equation $w = ax + b$ as follows

$$\mu_{Big}(x) = \frac{2}{1+\exp(-w)} - 1, \qquad x \in [0, +\infty)$$

$$\mu_{Sm}(x) = 2 - \frac{2}{1+\exp(-w)}, \qquad x \in [0, +\infty)$$

$$\mu_{Bigmod}(x) = \frac{2}{1+\exp(w)} - 1, \qquad x \in (-\infty, 0]$$

$$\mu_{Smmod}(x) = 2 - \frac{2}{1+\exp(w)}, \qquad x \in (-\infty, 0]$$

The presented below figures show the results of the classification of regions in one-dimensional space X using LLM.

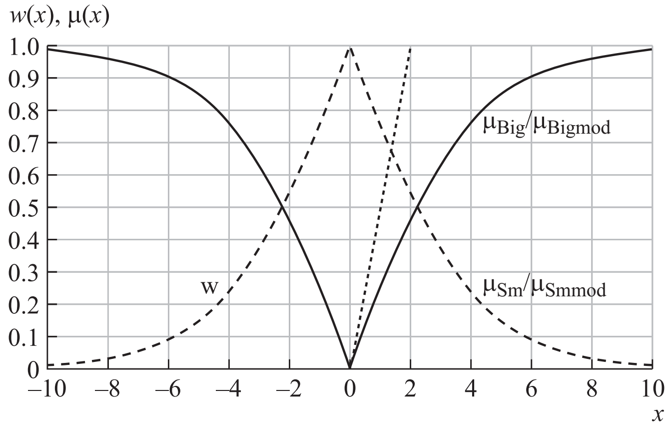

Fig. 1. Diagrams of membership function and logistic regression in log-model.

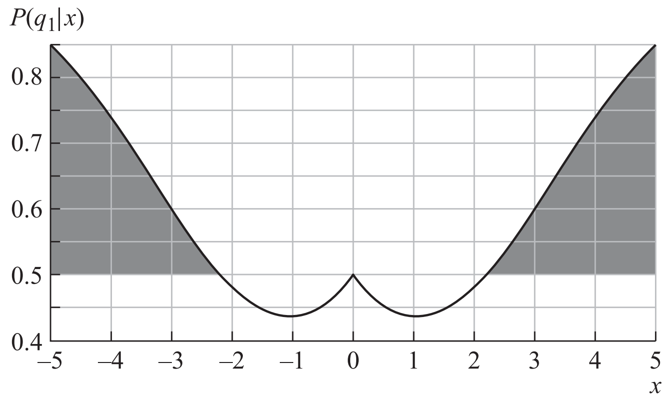

Fig. 2. Diagrams of intervals for class q_1 (painted) predicated on LLM.

7. SIMULATION EXPERIMENT

The developed approach to the analysis of predictors of complex technological processes is universal. At the same time, to ensure the possibility of its operation on real data and implementation in existing automation systems, an expert group of people with experience in the technological process is needed [13, 14]. Because the authors of this article have been working in the field of automation of railway transport management for more than 20 years and have the ability to create such an expert group, the proposed approach is considered in relation to ensuring the safety of the movement of cars on hump yards.

The goal of the strategy for the development of railways in the Russian Federation and other countries is to increase the share of railway transport in the overall structure of freight turnover. One way to solve this problem is to improve the planning and operation of marshalling stations [15, 16]. The most important and complex node of the marshalling station is the hump yard, where sorting of freight trains that have arrived at the station is carried out using the gravity rolling method. In the process of sorting, the train is pushed onto the hill, uncoupling the necessary groups of cars and rolling them down the hill under the action of gravity onto a given track by switches (Fig. 3).

To ensure the safety of the movement of cars in hump yard, monitoring and adjustment of the speed of groups of cars is necessary. It should be noted that the cutter movement speed adjustment is possible only in special areas, called braking positions [5].

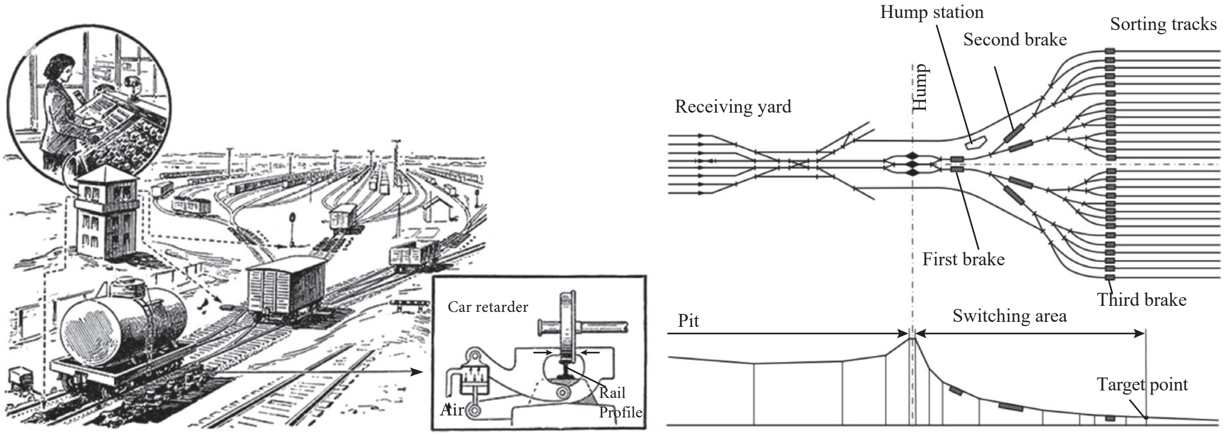

Fig. 3. Work of hump yard.

A dangerous (or abnormal) situation is considered to be the merging (collision) of cars between hills and sorting tracks. In this case, a collision can occur for various reasons:

—non-transfer of the switch (for example, the presence of more than one cut at the same time leads to the prohibition of switching the switch, which in turn leads to a collision),

—excessive braking force in the car retarder (for example, due to a malfunction of the retarder),

—when exceeding the speed of pushing cars up the hill.

At the same time, experts agree that the degree of danger of collision between hills and sorting tracks increases with the simultaneous presence of many single-car cuts with different weight categories [5, 17–19]. Therefore, the calculation of the hump parameters when designing new facilities or when analyzing situations of traffic safety violations is carried out for the most difficult case, when there are three or more single-car cuts with alternating running properties on the hump. From the above reasons for the increase in the degree of danger, it also follows that the complexity of the technological situation is affected by the distance between cuts, their speeds and a number of other parameters.

To assess the class affiliation of the technological situation, the following features were used:

- Driving properties of the car. Integral parameter (from 0 to 1), depending on particular characteristics of the cars (weight, length) and environmental parameters (wind, temperature, humidity). The formation of the parameter was carried out on the basis of the algorithms proposed in [20].
- Roll acceleration characteristic depending on the location of the car (m/s^2);
- Uncoupling speed (m/s);
- Distance to the separation point of the routes of cars (m) (According to [21], the combination of cars in curves is unacceptable, therefore, even if two cars go on the same track, cuts should be combined only in the sorting track).

A database of 5000 technological situations was used to train LLM. The experts group divided this database in two classes of complexity. The training data was collected from 20 sorting humps equipped with an integrated automatic control system of the sorting process (KSAU SP) [22, 23].

To assess the accuracy of the developed model, a test sample was used, containing 100 technological situations transferred to JSC NIIAS from JSC Russian Railways for the purpose of expert analysis. In all cases under consideration, manual intervention of the operator took place.

The result of manual intervention, shown on the screen of the electromechanical workstation, is shown in Fig. 4.

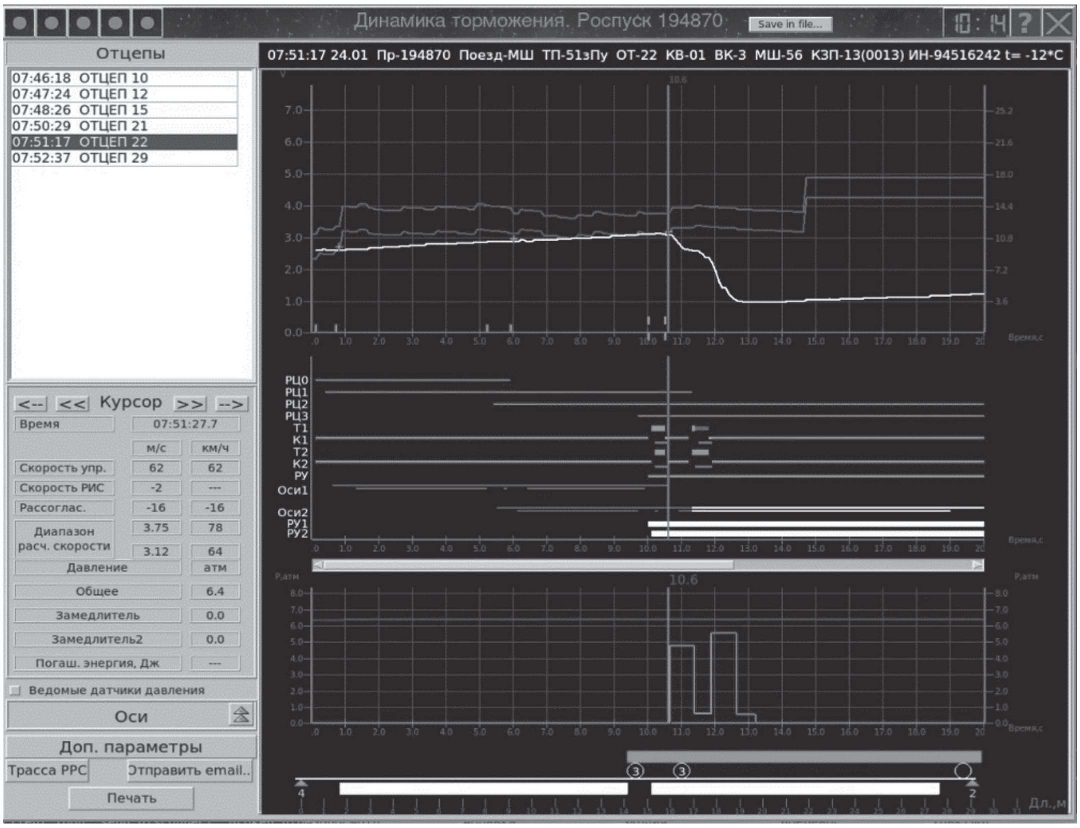

Fig. 4. The Example of manual interference.

The gray curve is the allowable car speed, the white curve is the current car speed. To the left of the blue vertical line, control was carried out by an automatic system. After that, the operator intervened in the process of movement, due to which the speed of the car dropped significantly.

An example of an emergency situation is shown in Fig. 5. In this example, cars 2, 3, 4, 5 and 6 were combined at the hill and were erroneously directed to the same route. An analysis of the described emergency situation showed that when the wagons were disbanded, the locomotive driver did not adhere to the recommended speed of pushing the wagons at the hill.

For each of 100 cases, state matrices were built:

$$X = \{\mathbf{x}_t\}. \tag{6}$$

For each of i parametrs $x_{it} \in \mathbf{x}_t$ the probability of occurrence of a difficult situation was estimated by formulas (2), (5), and for each state $\mathbf{x}_t$ the probability of occurrence of a difficult situation was estimated by DS-joins in the Zadeh basis:

$$P\left(q|\mathbf{x}_t\right) = \min_i P\left(q|x_i\right).$$

After that, the process of normalization of probabilistic indicators was carried out by bringing them to the interval $[0, 1]$. The technological situation was considered dangerous if in the matrix (6) there is element $\mathbf{x}_t$ with value $P(q|\mathbf{x}_t)$, exceeding 0.8.

As a result of the experiment, 21 complex TS were identified. In fact, among them, 19 situations were classified as complex. In 11 cases the permissible speed (5 km/h) was exceeded. In 7 cases, the failure of outdoor equipment was determined. In one case, over-braking of the car was revealed. The

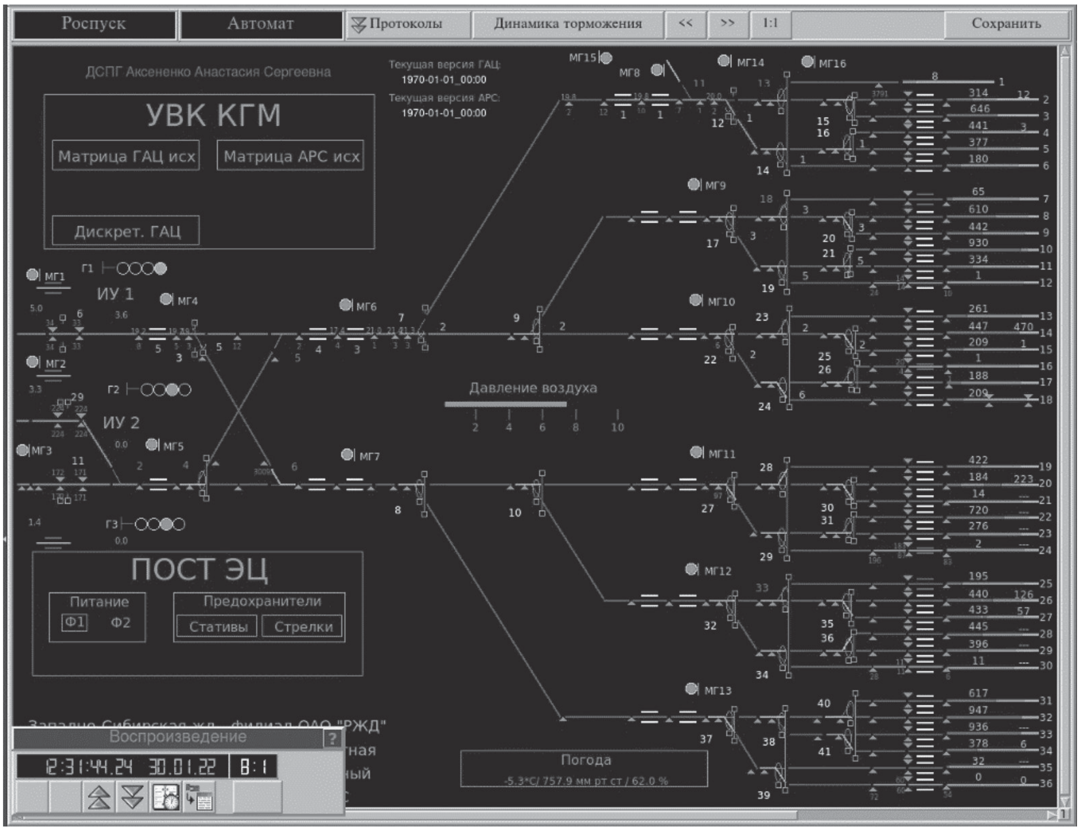

Fig. 5. The example of connection of several cars in hump yard.

remaining 2 cases are marked as falsely identified by the algorithm. Nevertheless, these situations were marked as borderline by employees of the signaling service of Russian Railways.

During the expert analysis of situations identified by the LLM as “simple”, 3 surges and 1 case of non-translation of the switch were detected, i.e., 4 false negative situations.

The accuracy assessment was carried out taking into account the calculation of the error matrix:

$$accuracy = \frac{\mathrm{TP} + \mathrm{TN}}{\mathrm{TP} + \mathrm{TN} + \mathrm{FP} + \mathrm{FN}},$$

where TP, TN, FP, FN—true-positive, true-negative, false-positive, false-negative objects (situations).

In the case of LLM for the situations described above, $\mathrm{TP} = 19$, $\mathrm{TN} = 96$, $\mathrm{FP} = 2$, $\mathrm{FN} = 4$ $accuracy = 0.95$, which indicates a high level of efficiency of the proposed approach.

8. CONCLUSIONS

The article presents a new predictive approach to analyzing states and predicting the behavior of complex technological processes based on predictive analytics technology. The proposed approach is based on the idea of detecting predictor events in a controlled process that precede the appearance of special classes of abnormal states of technological processes. The dependence of the predictor states of the technological process on the complexity of the technological situations that cause them has been established, which made it possible to approach the prediction of the states of the controlled process by estimating the complexity of the corresponding technological situations.

To implement the predictive approach, the authors proposed a hybrid logical-linguistic classifier model based on the combination of logistic regression, a probabilistic scheme for combining Dempster–Shafer evidence, and fuzzy Sugeno model. The inclusion of a fuzzy model in the hybrid classifier to assess the nature of the ES increases the expressiveness of the hybrid classifier and significantly expands its capabilities to identify and differentiate various classes of ES complexity. In particular, the logical-linguistic classifier allows, based on the analysis of the nature of the current technological situation, to identify various shades of the complexity of the ES by correcting the parameters of the logistic regression.

The conducted experiments on real data showed the effectiveness of the developed predictive model of the logical-linguistic classifier for assessing the states and predicting the behavior of technological processes occurring on hump yard. The generality of the proposed predictive approach and the universality of the developed classifier model make it possible to use them to solve a wide range of problems that arise in transport and other industrial production.

REFERENCES

1. Bukreev, V.G., Kolesnikova, S.I., and Yankovskaya, A.E., *Vyyavlenie zakonomernostei vo vremennykh ryadakh v zadachakh raspoznavaniya sostoyanii dinamicheskikh ob"ektov* (Identification of Patterns in Time Series in Problems of Recognition of the States of Dynamic Objects), Tomsk: Tomsk Polytechnic University, 2010.

2. Tsvetkov, V.Ya., Complex Technical Systems, *Obrazovatel'nye resursy i tekhnologii*, 2017, no. 3(20), pp. 86–92.

3. Vychuzhanin, V.V. and Vychuzhanin, A.V., Information Cognitive Simulation Model Of Complex Technical System, *Informatsionnye sistemy i tekhnologii IST-2020*, 2020, pp. 677–683.

4. Buravtsev, A.V. and Tsvetkov, V.Ya., Complicated Organizational and Calculating Systems, *Perspektivy nauki i obrazovaniya*, 2018, no. 4(34), pp. 293–300.

5. Shabelnikov, A.N. and Olgeyzer, I.A., Methods for Improving Safety in KSAU SP, *Avtomatika, svyaz', informatika*, 2017, no. 3, pp. 8–10.

6. Gurov, Y.V., Khatlamadzhiyan, A.E., Khilkov, D.V., and Shapovalova, Y., Adaptive Fuzzy Systems for Predictive Diagnostics of Railway Facilities, *Lecture Notes in Networks and Systems*, 2022, vol. 330 LNNS, pp. 170–179. https://doi.org/10.1007/978-3-030-87178-9_17

7. Sukhanov, A.V., Kovalev, S.M., Akperov, I.G., and Olgeyzer, I.A., Detection of Precursors of Bifurcations of a Dynamic System Based on the Analysis of the Structure of Its Fuzzy Model, *Integrirovannye modeli i myagkie vychisleniya v iskusstvennom intellekte IMMV-2020: Sbornik nauchnykh trudov XI Mezhdunarodnoi nauchno-prakticheskoi konferentsii* (Integrated Models and Soft Computing in Artificial Intelligence IMMV-2022: Collection scientific works of the XI International Scientific and Practical Conference. In 2 volumes, Kolomna, May 16–19, 2022. Volume 1.—Kolomna: All-Russian public organization "Russian Association of Artificial Intelligence," 2022, pp. 137–144.

8. Gorrini, V., Salome, T., and Bersini, H., Self-Structuring Fuzzy Systems for Function Approximation, *Proceedings of 1995 IEEE International Conference on Fuzzy Systems. IEEE*, 1995, T. 2, pp. 919–926.

9. Quost, B., Masson, M.-H., and Denœux, T., Classifier Fusion in the Dempster-Shafer Framework Using Optimized t-Norm Based Combination Rules, *International Journal of Approximate Reasoning*, 2011, no. 52(3), pp. 353–374.

10. Denoeux, T., Logistic Regression Revisited: Belief Function Analysis, *International Conference on Belief Functions*, Springer, Cham, 2018, pp. 57–64.

11. Dempster, A.P., Upper and Lower Probabilities Induced by a Multivalued Mapping, *Annals of Mathematical Statistics*, 1967, no. 38, pp. 325–339.

12. Yager, R.R., Measures of Entropy and Fuzziness Related to Aggregation Operators, *Information Sciences*, 1995, T. 82, no. 3–4, pp. 147–166.

13. Afanasieva, T.V., Granulation of Multivariate Time Series in The Problem of Descriptive Analysis of the State and Behavior of Complex Objects, *Automation and Remote Control*, 2022, vol. 83, no. 6, pp. 884–893.

14. Trofimov, V.B., An Approach to Intelligent Control of Complex Industrial Processes: An Example of Ferrous Metal Industry, *Automation and Remote Control*, 2020, vol. 81, no. 10, pp. 1856–1864.

15. Kazantseva, L.S. and Yugrina, O.P., Normalization of Shipment Delivery Dates and Technology of Transportation Process, *Byulleten' transportnoi informatsii*, 2015, no. 6, pp. 29–33. ISSN 2072-8115

16. Pokrovskaya, O.D., Logistics Transport Systems of Russia in the Context of New Sanctions, *Byulleten' rezul'tatov nauchnykh issledovanii*, 2022, no. 1, pp. 80–94.

17. Mukha, Yu.A., Tishkov, L.B., Sheykin, V.P., et al., *Posobie po primeneniyu pravil i norm proektirovaniya sortirovochnykh ustroistv* (Manual on the Application of Rules and Norms for Design of Assorting Works), Moscow: Transport, 1994. https://doi.org/10.1051/matecconf/201821602012

18. *Pravila i normy proektirovaniya sortirovochnykh ustroistv na zheleznykh dorogakh kolei 1520 mm* (Rules and Design Standards for Assorting Works on Railways with Track 1520 mm), Moscow: Tehinform, 2003.

19. Bessonenko, S.A., *Doctoral thesis*, Moscow: Moscow State University of Railway Engineering, 2011.

20. Olgeyzer, I.A., Sukhanov, A.V., Shabelnikov, A.N., and Ignatieva, O.V., Fuzzy Approach to Car Retarding Adaptation on Hump Yards, *Lecture Notes in Networks and Systems*, 2022, vol. 330 LNNS, pp. 161–169. DOI 10.1007/978-3-030-87178-9_16

21. Rules for the technical operation of the railways of the Russian Federation, approved by Order of the Ministry of Transport of Russia dated July 23, 2022, no. 250.

22. Utility model patent no. 95623 U1 Russian Federation, IPC B61L 17/00. Integrated automation system for sorting process control (KSAU SP): no. 2010109685/22: Appl. 03/15/2010: publ. 07.10.2010 / Danshin, A.I., Zolotarev, Yu.F., Odikadze, V.R., et al., applicant JSC "Scientific Research and Design Institute of Informatization, Automation and Communication in Railway Transport."

23. Andronov, D.V., Experience in Operating the KSAU SP System, *Avtomatika, svyaz', informatika*, 2013, no. 11, pp. 16–18.

This paper was recommended for publication by M.F. Karavai, a member of the Editorial Board