Identification and Classification of Train Timetable Deviations Using OSTIS Technology

Aleksandr Erofeev
Belarusian State University of Transport
Gomel, Belarus
erofeev_aa@bsut.by

Ilya Erofeev
Belarussian State University of Informatics and Radioelectronics
Minsk, Belarus
ilerofv@gmail.com

***Abstract*—A posteriori model of dispatcher train schedule correction is described, an integrated scheme of a hybrid operational problem solver is established, an algorithm for inducing a tree for classifying operational situations in TS is proposed, a method for identifying deviations in the train schedule using OSTIS technologies is described, the basic concepts of the subject area "Dispatching train schedule correction" are formulated and described**

***Keywords*—Intelligent transportation process management system, train timetable, train schedule, transportation process, train, station, station interval, ontology, OSTIS technology**

Introduction

Intelligent systems play a key role in the modern world, providing effective management of complex processes and tasks. In particular, in the field of railway logistics and transport, the development of innovative approaches to the management of the transportation process is becoming an increasingly urgent task. One of the important aspects here is the effective planning and adjustment of train schedules, which makes it possible to optimize the use of infrastructure and resources. Within the framework of this paper, a posteriori model of dispatching train schedule adjustments will be described, an integrated scheme of a hybrid operational problem solver is established, an algorithm for inducing a tree for classifying operational situations in TS is proposed, a method for identifying deviations in the train schedule using OSTIS technologies is described, and the basic concepts of the subject area "Dispatcher Train schedule adjustment" are formulated and described

I. Task definition

The train schedule (TS) is the basis for the organization of the transportation process on the railway. Its development and subsequent implementation make it possible to link the operation of railway stations, sections, operation of locomotives, etc. into a single technology. However, in an operational situation, a number of situations arise that do not allow the regulatory train schedule to be implemented: late departure of the train from the station due to the need to perform additional operations during maintenance of the trains, an increase in the parking time of the passenger train at an intermediate station to organize the boarding and disembarkation of a large number of passengers, etc. A dispatch control system is designed to respond to such situations on the railway.

However, in practice, due to limited time and information resources, the dispatcher may not always be able to correctly develop an option for dispatching adjustments. In this regard, it is proposed to create a subsystem for intelligent dispatch control of the TS within the framework of the intelligent transportation process management system (ITPMS). [1]

The a posteriori model of intelligent dispatch correction of the TS assumes the solution of the problem **situational modeling** [2]:

- identification of deviations of the actual train line from the forecast one;
- classification;
- subsequent liquidation by one of the available methods.

It is assumed that at the stage of development of the regulatory TS preliminary optimization of the schedule was carried out and, taking it into account, control solutions have already been formed in the adjacent subsystems of the ITPMS. As a criterion of effectiveness in the a posteriori model, it is proposed to use the conditions of minimal changes in the TS relative to the normative one.

In this case, the hybrid solver of the operational task (HSOT) of the control correction of the TS will include procedures for identifying deviations, selecting a scenario for introduction the train into the TS and implementing the control correction in accordance with the selected action algorithm (Figure 1) [3].

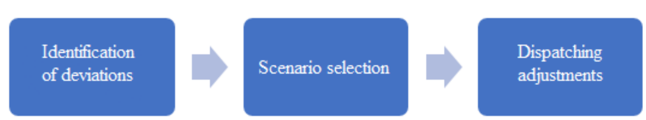

Figure 1. Enlarged structure of the HSOT "Dispatch correction of the TS"

ITPMS should provide for the possibility of training, i. e. the formulation of new control management systems, new adjustment scenarios and algorithms for adjusting the TS depending on the initial scenarios.

In order to organize the set of possible options for dispatching adjustments and choosing a rational one, it is proposed to form a decision tree of the TS. A decision tree is a method of representing decision rules in a hierarchical structure consisting of two types of elements – nodes and leaves. The decision rules are located in the nodes and the examples are checked for compliance with this rule by some attribute of the training set [4].

In general, the *subject area* for solving the problem under consideration will include possible types of dispatcher adjustments, including [5]:

- changing the order of crossing trains;
- changing the train's receiving path due to the discrepancy between the length of the station tracks and the length of the train, if necessary, the train must stop at the station that it must pass without stopping according to the TS;
- selection of available tracks for receiving a train if it is necessary to stop at a station that, according to the TS, it must pass without stopping;
- organization of non-stop passage by train at the station where, according to the TS, the train has a stop;
- changing the arrival, departure, and passage times of passenger trains;
- canceling the TS route option;
- laying an additional thread in the TS;
- allocation of time in the TS for the production of works in the "window", etc.;
- change of the time for the provision of "windows", etc.

A lot of dispatcher adjustments to the TS can be reduced to five possible scenarios for the formation of MD:

- adjustment of the TS thread with subsequent TS changes;
- change of the train departure time from the station.
- linking the train to another nearby free route on the schedule;
- development of an additional route taking into account the preservation of the timetable of other trains;
- canceling the TS option.

During the operation of the ITPMS, other types of dispatching adjustments, scenarios for the formation of MD and methods of adjustments can be identified. That is, the system of dispatching adjustments should be developing and learning. For these purposes, it is proposed to use inductive rules, and the learning process itself will be the induction of trees.

The TS decision tree is a model based on teacher-led learning, and target values should be set in the training set of dispatcher adjustments. At the same time, since the target variables are discrete and can be described by logical relationships (class labels), the model will be a classification tree [3].

The key feature of intelligent dispatch correction of the TS that, unlike classification models used in other subject areas, in which a value or function is considered as a leaf, it is proposed to consider the algorithm for performing actions for dispatching correction of the TS as a leaf of the tree.

Then HSOT will include the following operational tasks:

- definition of a set of class labels for each operational situation;
- selection using the classification tree of a rational algorithm for dispatching TS adjustment;
- implementation of the dispatcher adjustment algorithm for the current operational situation.

II. Algorithm for the induction of a classification tree for operational situations in TS.

Let's set up a training set of operational situations *S*, (discrepancies between the planned and actual TSs), containing *n* options, for each of which a class label is set $C_i(i = 1...k)$, and *m* attributes $A_j(j = 1...m)$, which determine whether an object belongs to a particular class. At the same time, an automatic dispatch correction algorithm is installed for each C_i class.

The purpose of TS tree induction is to determine for any operational situation A_j class C_i (dispatcher adjustment algorithm).

There are three possible cases.

1) All examples of multiple operational situations *S* have the same class label C_i (i.e. all training examples belong to only one class). Obviously, learning in this case does not make sense, since all the examples presented to the model will be of the same class, which will "learn" to recognize the model. In this case, the decision tree itself will be a leaf associated with the class C_i. The practical use of such a tree is meaningless, since it will assign any new object only to this class.
2) The set *S* does not contain any examples at all, i.e. it is an empty set. In this case, a leaf will also be created for it (it makes no sense to apply the rule to create a node to an empty set), the class of which will be selected from another set (for example, the class that is most often found in the parent set).
3) Set *S* contains training examples of all classes C_k. In this case, you need to split the set *S* into subsets associated with classes. To do this, select one of the attributes A_j sets *S*, which contains two or more unique values (a_1, a_2, ..., a_p), where *p* is

the number of unique attribute values. Then the set of S is divided into p subsets (S_1, S_2, ..., S_p), each of which includes examples containing the corresponding the attribute value. Then the next attribute is selected and the partitioning is repeated. This procedure will be repeated recursively until all the examples in the resulting subsets are of the same class.

When using this technique, the decision tree will be built from top to bottom (from the root node to the leaves). Currently, a significant number of decision tree learning algorithms have been developed: ID3, CART, C4.5, C5.0, NewId, ITrule, CHAID, CN2, etc. [6]. In this paper, we do not evaluate the effectiveness of using one algorithm or another, but consider the specifics of setting the problem of induction of a decision tree for a dispatch control system.

A distinctive feature of TS from traditional classification objects is its continuous change over time. In this regard, it is proposed to consider a variety of parameters that characterize the TS threads as temporal data. Then the task of generalizing and classifying operational situations in the presence of temporal data is formulated as follows.

Let's assume that the graph of completed movement (GCM) is based on data on changes in the states of block sections q when trains pass through them. The states change at some discrete points in time: $t = 0, 1, 2, 3, \ldots$. Then the train situation at the control range at some point in time i can be represented as a vector

$$S_i =< x_1(t = i), x_2(t = i), \ldots, x_q(t = i), \\ t = 1 > 1. \quad (1)$$

In order to trace the dynamics of the trains' progress along the section, their possible deviations from the planned lines of the TS, and the dynamics of changes in deviations (lagging or overtaking), it is necessary to consider an ordered set of such vectors obtained over a finite time interval t_i, t_{i+r-1}), $r > 1$.

Let q parameters be considered for the time interval om of length r. Let's present such data in the form of a matrix (Table I).

Each cell of the matrix represents the value of the location of the train q at the time i. Then each column of the matrix will describe a thread in the graph of the executed movement, and each row will represent the train position on the section at time points, respectively $i, i + 1, \ldots, i + r - 1$.

However, the operational situation on the stage is characterized not only by the relative location of trains on the section, but also by their compliance with the planned TS. In this regard, it is proposed to supplement the matrix with train values of the current TS (CTS).

Based on the data on the GCM and CTS trains, an estimated train situation is formed (the location of each

Table I
Dynamic object of generalization of the graph of completed movement

	Train 1	Train 2	...	Train g	Time t
(S_i)	$x_1(t = i)$	$x_2(t = i)$		$x_g(t = i)$	i
(S_{i+1})	$x_1(t = i+1)$	$x_2(t = i+1)$		$x_g(t = i+1)$	$i+1$
(S_{i+2})	$x_1(t = i+2)$	$x_2(t = i+2)$		$x_g(t = i+2)$	$i+2$
...					
(S_{i+r-1})	$x_1(t = i+r-1)$	$x_2(t = i+r-1)$		$x_g(t = i+r-1)$	$i+1$

of the trains in motion at each moment of the planned period). The estimated value of the train location at the time t_{i+r+v} is defined as

$$x_v(t = i + v + r) = x_q(t = i + v) + \\ \Delta x^{\text{прогноз}}_{r-q}(\Delta t = r - q). \quad (2)$$

Then the operational situation at the landfill will be characterized by a set of arrays of information about executed, planned and calculated train conditions (Table II).

Table II
Dynamic Operational Situation Generalization Object (CTS)

	Train 1	Train 1 (schedule)	Train 2	Train 2 (schedule)	...	Train q	Train q (schedule)	Time (t)
S_i	GCM	Schedule						
S_{i+1}								
S_{i+2}								
...								
S_{i+v}								
...	Forecast							
S_{i+r}								

Let's call the structure presented in Table II a dynamic object of generalization of the operational situation. Then the task is to assign the appropriate operational situation to a specific class, depending on the methods of dispatching adjustments being implemented.

The assignment of operational environments to the appropriate classes is performed using a decision tree.

Formally, a decision tree is a weighted directed graph $T = (V, E)$. In the set of vertices V, select the vertex $v_0 \in V$ – the root of the tree. We will divide all the vertices into two classes:

- $V_i \subset V$ – the set of internal vertices (nodes) of the tree; V_i includes such vertices, of which arcs come out;
- $V_l \subset V$ – the set of external, final, tree vertices (leaves); V_l includes it includes vertices from which arcs do not exit; V_i and V_l form a partition of the set of vertices of the V decision T:

$$V_i \cap V_l = \emptyset; V_i \cup V_l = V.$$

The internal vertices of the V_i tree correspond to the attributes that are used in classifying the operational environment. The leaf vertices V_l correspond to the algorithms for dispatcher adjustment of TS.

Each arc of the e decision tree is weighted by the condition "attribute = attribute value" (for qualitative attribute values) or "attribute σ attribute value" (for quantitative attribute values, $\sigma \in \{\geq>=\}$).

III. The attribute selection rule for dividing a set of operational situations into subsets

The general rule for choosing an attribute for dividing a set of operational situations into subsets can be formulated as follows: the selected attribute should split the set of observations in the node so that the resulting subsets contain examples with the same class labels, or are as close to it as possible, i. e. the number of objects from other classes ("impurities") in each of these sets was as little as possible. To select attributes for dividing a set of operational environments into subsets, it is proposed to use an information-theoretical criterion based on the concept of information entropy.

$$H = -\sum_{i=1}^{n} \frac{N_i}{N} \log(\frac{N_i}{N}), \quad (3)$$

where n is the number of classes in the original subset; N_i – the number of examples of the ith class; N – the total number of examples in the subset.

Entropy is considered as a measure of the heterogeneity of a subset according to the classes represented in it. When classes are presented in equal proportions and classification uncertainty is greatest, entropy is also maximal. If all the examples in the node belong to the same class, i. e. $N = N_i$, the logarithm of one turns the entropy to zero.

Thus, the best attribute of splitting A_j will be the one that ensures the maximum reduction in the entropy of the resulting subset relative to the parent. In practice, however, they do not talk about entropy, but about the inverse of it, which is called information. Then the best partitioning attribute will be the one that will maximize the information gain of the resulting node relative to the original one.:

$$Gain(A) = Info(S) - Info(S_A), \quad (4)$$

where $Info(S)$ – information related to the subset of S before splitting; $Info(S_A)$ – information related to the subset obtained by splitting by attribute A.

As a result of the conducted research, it has been established that when classifying operational situations and selecting algorithms for dispatching adjustments to TS, it is advisable to use the following attributes A_j:

- for each of the trains, the estimated arrival time of the train at the station corresponds to the planned value;
- for each of the trains, the estimated departure time of the train from the station corresponds to the planned value;
- for each of the trains, the correspondence of the stage running times to the planned values;
- for each of the pairs of trains located next to each other in the TSTHE observance of station intervals;
- for each of the pairs of trains located next to each other in the TSTHE observance of inter-train intervals;
- availability of reserves in the driving times;
- availability of reserves in station and inter-train intervals;
- availability of reserves during train stops;
- availability of backup threads in TS;
- the category of the train that deviated from the planned TS (passenger, suburban, freight, long-range, heavy, etc.).

To implement the induction algorithm for the classification tree of operational situations in TS, the key task is the correct ***identification of the current operational situation***. It is proposed to use OSTIS technologies for this purpose [7]–[10].

IV. The ontology of the TS domain.

The main objects of the TS subject area are stations and trains. Such concepts as stages, station tracks, schedules and intervals are also highlighted to describe them (Figure 2).

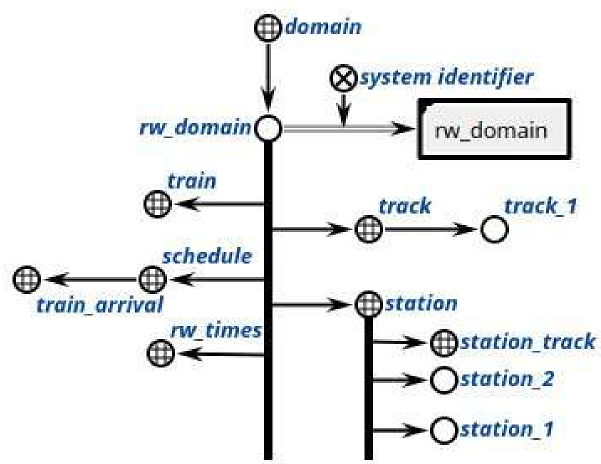

Figure 2. Objects of the TS subject area

In turn, trains have their own characteristics and are divided into freight and passenger trains. Passenger trains are generally divided into international, interregional (standart), regional, and urban trains. This classification is extensible. Freight or Catgo trains are divided into accelerated, through, district and collected, and may also have additional properties, such as container, heavy, long-composite, etc. An example of defining a train according to these classifications is *train_1* (Figure 3).

There are stages adjacent to the station. For each of the runs, the length, the value of station and inter-train

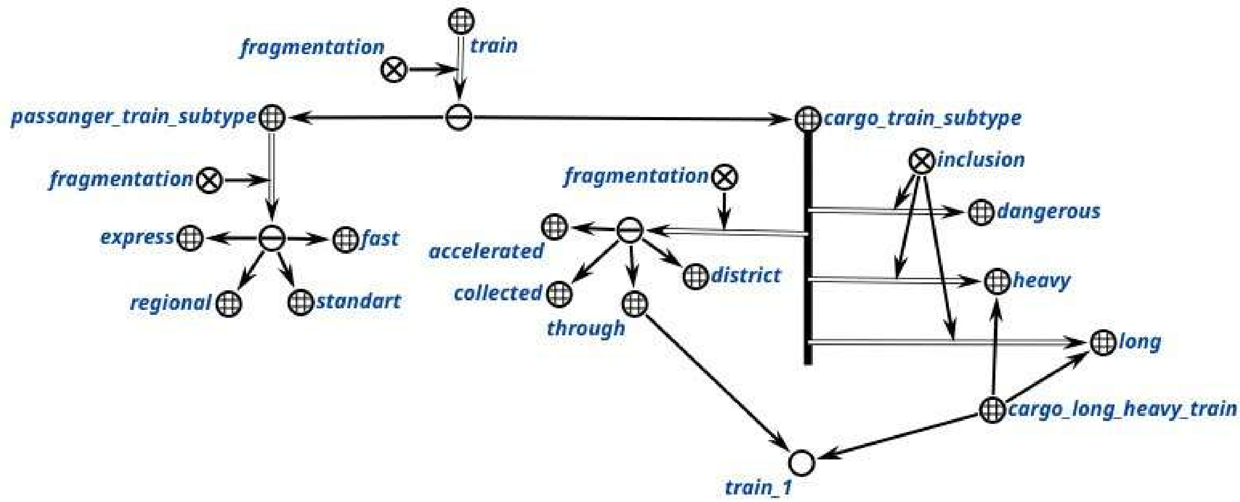

Figure 3. TS subject area section "Classification of trains"

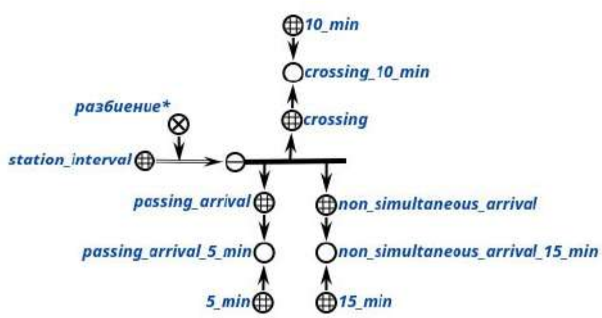

Figure 4. General view of the interval

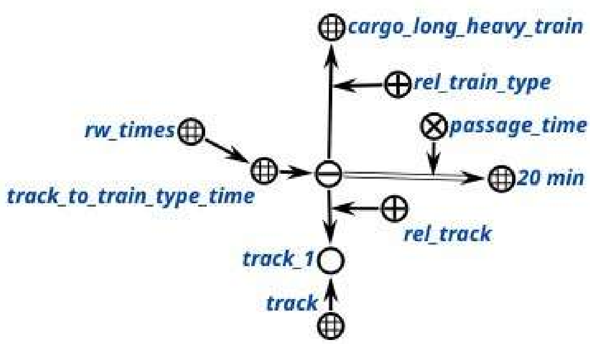

Figure 5. Travel time for the stage

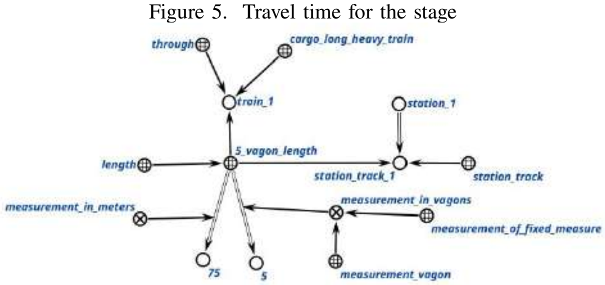

Figure 6. Description of train length and station track length

intervals for all combinations of different categories of trains should be determined (Figure 4), as well as the travel time for different types of trains (Figure 5).

In addition, for each of the crossings, the parameters of the approaches to the station should be determined, including the slopes in the areas of approach to the station (Figure 6).

Depending on the magnitude of the slope on the approach to the station and the characteristics of the train, a decision is made on the possibility or impossibility of stopping the train in question at the station.

Each station includes in its structure specialized tracks designed for receiving and departing trains. The static characteristics of the tracks are their length (capacity in wagons), specialization, the possibility of receiving specialized trains, etc.) (Figure 7).

Each track can be free or occupied by a specific train at any given time.

In addition to the infrastructure and characteristics of the trains, it is necessary to describe the regulatory TS (Figure 8).

The regulatory TS defines for each train the arrival and departure times of trains at each station of the railway section, as well as the occupation of tracks by operations of passing, stopping and parking.

This ontology is the basis for the construction of the HSOT "Dispatching adjustment of TS". Of course, this ontology is not exhaustive and will be refined, modified and supplemented as it is used.

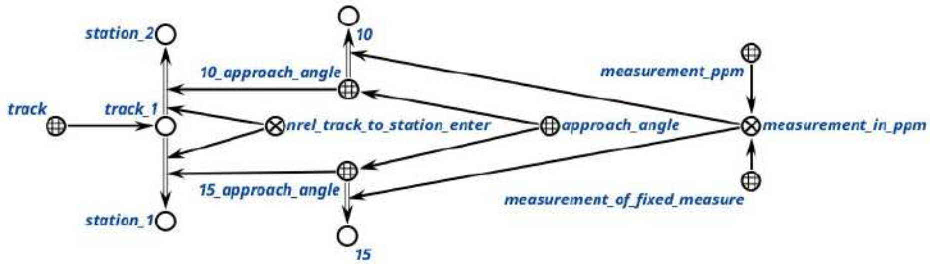

Figure 7. Description of the track

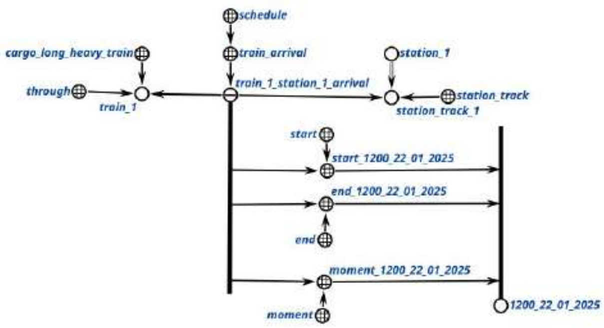

Figure 8. Description of the regulatory TS the station

V. Conclusion

In conclusion, it is worth noting that the development of intelligent transportation process management systems based on OSTIS technologies opens up new prospects for optimizing railway logistics and improving the efficiency of transport processes. The a posterior model of dispatching train timetable adjustments described in this paper represents an important step towards automation and intellectualization of railway transport management.

The use of classification tree induction algorithms to analyze and adjust operational situations in train schedules opens up new opportunities for rapid response to changing conditions and provides more flexible and efficient railway traffic management. The further development and application of such innovative approaches in the railway industry has the potential to change and improve transportation processes, providing a higher level of service and resource optimization.

References

[1] Erofeev A. A. Intelectualnaya sistema upravlenia perevozochnim processom [Intelligent control system for the transportation process in railway transport]. The Ministry of Education of the Republic. Belarus, Belarusian state University of Transport, BSUT, Gomel, 2022, 437 p.

[2] Pospelov D. A. Situacionnoe upravlenie: teoriya i praktika. [Situational management: theory and practice], Moskow, Science, 1986, 288 p.

[3] Erofeev A. A. Aposteriornaya model' intellektual'noj dispetcherskoj korrektirovki grafika dvizheniya poezdov [A posterior model of intelligent dispatching train schedule adjustments]. Problems of railway stations and junctions perspective development: International collection of scientific papers, Gomel:BSUT, 2021, № 3, pp. 96-107.

[4] Quinlan J. R. Induction of decision trees. Machine Learning, № 1(1), P. 81-106.

[5] Erofeev A. A., Erofeeva H. A. Avtomatizirovannaya sistema razrabotki prognoznogo grafika dvizheniya poezdov na Belorusskoj zheleznoj doroge [Automated system for the development of forecast train schedules on the Belarusian Railway]. Bulletin of BSUT: science and transport, 2011, № 1 (22), pp. 59–64.

[6] Murthy S. Automatic construction of decision trees from data: A Multi-disciplinary survey, Boston, 1997, 49 p.

[7] V. V. Golenkov, "Ontology-based design of intelligent systems," in Open semantic technologies for intelligent systems, Iss. 1, Minsk, BSUIR, 2017, pp. 37–56.

[8] V. Golenkov and N. Guliakina, "Principles of building mass semantic technology component design of intelligent systems," in *Open semantic technologiesfor intelligent systems*, Minsk, BSUIR, 2011, pp. 21–58.

[9] Shunkevich D. V. Agentno-orientirovannye reshateli zadach intellektual'nyh sistem: dis. [Agent-based problem solvers of intelligent systems: dis] … Doctor of Technical Sciences, Minsk, 2017.

[10] V. V. Golenkov, Tehnologija kompleksnoj podderzhki zhiznennogo cikla semanticheski sovmestimyh intellektual'nyh komp'juternyh sistem novogo pokolenija [Technology of complex life cycle support of semantically compatible intelligent computer systems of new generation], Minsk, Bestprint, 2023, 1063 p.

ИДЕНТИФИКАЦИЯ И КЛАССИФИКАЦИЯ ОТКЛОНЕНИЙ В ГРАФИКЕ ДВИЖЕНИЯ ПОЕЗДОВ С ИСПОЛЬЗОВАНИЕМ ТЕХНОЛОГИЙ OSTIS

Ерофеев А. А., Ерофеев И. А.

Дано описание апостериорной модели диспетчерской корректировки графика движения поездов, установлена укрупненная схема гибридного решателя эксплуатационных задач, предложен алгоритм индукции дерева классификации эксплуатационных ситуаций в ГДП, описан метод идентификации отклонений в графике движения поездов с использованием технологий OSTIS, сформулированы и описаны основные понятия предметной области «Диспетчерская корректировка графика движения поездов»

Received 31.03.2025