Data and Knowledge Base on the Basis of the Expanded Matrix Model of Their Representation for the Intelligent System of Road-Climatic Zoning of Territories

A Yankovskaya[1, 2, 3, 4, 5, a], D Cherepanov[1, b], O Selivanikova[3, c]

[1]Tomsk State University of Architecture and Building, Russia, 634003, Tomsk, 2 Solyanaya Sq. Tel. (3822) -65-07-82
[2]Tomsk State University, Russia, 634050, Tomsk, 36 Lenin Ave.
[3]Tomsk Polytechnic University, Russia, 634050, Tomsk, 2 Lenin Ave.,
[4]Tomsk State University of Control Systems and Radioelectronics, Russia, 634050, Tomsk, 40 Lenin Ave.
[5]Siberian State Medical University, Russia, 634050, Tomsk, 2 Moscovskiy Trakt

E-mail: [a]ayyankov@gmail.com, [b] d_n_ch@mail.ru, [c] selov@tpu.ru

Abstract. An extended matrix model of data and knowledge representation on the investigated area, as well as a matrix model of data representation on the territory under investigation, are proposed for the intelligent system of road-climatic zoning of territories (RCZT) – the main information technology of RCZT. A part of the West Siberian region has been selected as the investigated territory. The extended matrix model of knowledge representation is filled out by knowledge engineers with participation of highly qualified experts in the field of RCZT. The matrix model of data representation on the territory under investigation is filled out by persons concerned in RCZT of the motor-roads management system.

1. Introduction

The existing motor-roads network in the Russian Federation currently does not allow to fully meet the needs of the economy and the competitiveness of the international transportation of goods by motor-road throughout the territory of Russia. The importance of motor-road transportation cannot be overestimated in preservation the territorial integrity of the Russian Federation, its geopolitical influence, and its competitiveness. [1] In this connection, the development of modern information technologies in the motor-road management system assumes a special prominence.

Road-climatic zoning of territories (RCZT) includes the allocation on the territory of the Russian Federation of road-climatic zones characterized by one or another combination of heat and moisture, which stipulates, within the limits of a single zone, the formation of interrelated types of soil and vegetation [2 – 7]. However, other factors that significantly affect the condition of motor-roads can vary significantly within each zone [3 – 5]. In this regard, approaches to RCZT have to be refined.

Consideration of specificities related to natural-and-climatic conditions of West Siberian and other regions of Russia, in addition to the existing design standards in the motor-road industry [2], allows to divide road-climatic zones into homogenous territories [3 – 5] characterized by typical and peculiar

geocomplexes (climate, geology, topography, and other conditions). Within each homogeneous territory the homotypic road structures, in the first instance especially roadbed and pavement design, are characterized by approximately the same strength and stability. Such division into homogeneous territories should prevent significant financial and economic losses throughout the life cycle of motor-roads, associated with their maintenance to the desired safety standards. In addition, it is necessary to develop a modern methodology of RCZT for clarification of the boundaries of homogeneous territories that had been allocated on the basis of previously used approaches. Thus, the development of new information technologies in the field of RCZT is beyond question.

The proposed in the publication [8] intelligent information technology of road-climatic zoning of territories (IIT RCZT) will allow to attribute the area where a motor-road is planned to be built to a known type of the homogeneous territory. For this purpose, it is necessary to use matrix way of data and knowledge representation in the intelligent system (IS) of RCZT, which serves as the basis of IIT RCZT consisting of an expanded matrix model of data and knowledge representation on the investigated area of the West Siberian region and a matrix model of data representation on the region under investigation, as described is in the publication [8].

Below is given the matrix way of data and knowledge representation used for RCZT; structuring of data and knowledge; characteristic and classification features and their values; formation of the database and knowledgebase in IS RCZT.

2. Matrix way of data and knowledge representation

In the publication [8] it is proposed to use an extended matrix model of data and knowledge representation for RCZT. Integer values of characteristic features, including group features (features each of which is divided into another features that take integer values corresponding to certain intervals) as well as compulsory features are elements of the expanded description matrix ($\mathbf{Q^e}$)..

Integer matrix of descriptions ($\mathbf{Q^e}$) defines a description of objects in the space of characteristic features $z_1, z_2, \ldots, z_m$. The column of the matrix $\mathbf{Q^e}$ corresponds to each characteristic feature. The row of the matrix $\mathbf{Q^e}$ corresponds to each reference point at which the values of characteristic features are defined. Thus, the value of an integer characteristic feature (including the compulsory feature corresponding to one of the reference points) is the element of the matrix $\mathbf{Q^e}$ [8]. It shall be noted that compulsory features are not involved in the definition of regularities, but serve only to map the position of zones, subzones, and road districts.

Moreover, the description matrix $\mathbf{Q^e}$ is complemented by rows that represent the knowledge of highly qualified experts. It shall be noted that if the value of the characteristic feature in the additional row is not known or not indicated by highly qualified experts, this fact is designated by a dash ("–") in the respective element of the matrix $\mathbf{Q^e}$.

$$
\mathbf{Q^e}=
\begin{array}{c}
\begin{array}{ccccccc}
z_1 & z_2 & z_3 & z_4 & z_5 & z_6 & z_7
\end{array}\\
\left[\begin{array}{ccccccc}
3 & 4 & 8 & 9 & 5 & 1 & 3\\
2 & 5 & 7 & 9 & 4 & 1 & 5\\
3 & 5 & 7 & 8 & 3 & 0 & 8\\
4 & 2 & 6 & 8 & 4 & 1 & 5\\
3 & 3 & 8 & 8 & 4 & 1 & 5\\
2 & 3 & 8 & 8 & 4 & 1 & 3\\
1 & 2 & 8 & 8 & 3 & 2 & 3\\
1 & - & 7 & 8 & 4 & 0 & 6\\
2 & 6 & 7 & 9 & 5 & 1 & 3\\
1 & 4 & 8 & 8 & 3 & 0 & 4
\end{array}\right]
\begin{array}{c}
1\\2\\3\\4\\5\\6\\7\\8\\9\\10
\end{array}
\end{array}
\qquad
\mathbf{R^e}=
\begin{array}{c}
\begin{array}{ccc}
k_1 & k_2 & k_2
\end{array}\\
\left[\begin{array}{ccc}
1 & 1 & 1\\
1 & 1 & 2\\
1 & 2 & 1\\
1 & 2 & 1\\
1 & 1 & 1\\
2 & 1 & 1\\
1 & 2 & 2\\
2 & 2 & 1\\
2 & 1 & 2\\
2 & 2 & 2
\end{array}\right]
\end{array}
$$

Fig. 1 Matrices $\mathbf{Q^e}$ and $\mathbf{R^e}$

Integer matrix of distinctions ($\mathbf{R}^e$) defines the partition of objects into equivalence classes for each classification mechanism. Integer values of classification features are elements of the distinctions matrix $\mathbf{R}^e$. We shall restrict ourselves to the distinctions matrix of the diagnostic type where each subsequent column parts the previous one into equivalence classes. In accordance with the practice of RCZT [3 – 5] we shall use three classification features of the diagnostic type, which correspond to: 1st – zones, 2nd – subzones, 3d – road districts.

Some additional rows of the distinctions matrix, filled out by highly qualified experts, may differ from the previously filled on the basis of experimental data, because in the learning sample some combinations of values of classification features can be absent.

Information on the objects under investigation is represented only by the matrix $\mathbf{Q'}^e$, similar to the matrix $\mathbf{Q}^e$, and determined the description of the territory under investigation subjected to RCZT.

An example of a description and distinctions extended matrices for ten reference points is given on fig. 1. Rows formed on the basis of the knowledge of highly-qualified experts are absent in matrices $\mathbf{Q}^e$, $\mathbf{R}^e$.

3. Structuring data and knowledge

Formation of the data and knowledge base is carried out on the basis of structurization data and knowledge given in the article [8]. To create IIT RCZT it is advisable to use the following three types of characteristic features proposed in [8]: characteristic features (CF), group CF, and compulsory CF.

We shall list 6 CF: type of vegetation; terrain relief; calculated soil moisture (CSM); content of sand particles; evaporation from the surface of the roadbed in summer months; type (type and subtype) of soil.

The following are 7 group CF: Selyaninov hydrothermal coefficient (SHC); plasticity index (PI); size of soil particles (SSP); average annual air temperature (AAT); snow cover depth (SCD); soil freezing depth (SFD); relative frost heave (RFH).

The following four features are used as compulsory features that do not participate in the identification of particularities but serve only to map the position of road-climatic zones, subzones, and road areas: name of the zone; name of the subzone; name of the road district; name of the reference point.

In addition to characteristic features there are three classification features: zone; subzone; road district.

4. Characteristic and classification features and their values

The knowledge base is formed on the basis of description and distinctions matrices [9]. The values of elements of the description matrix are determined based on the information database. The values of characteristic features or intervals or their values are encoded by integers. Intervals of partitioning of feature values are determined on the base of expert knowledge. The result is a set of integer values of characteristic features.

The following are integer values encoding characteristic features, associated to columns of the description matrix:

z1 – vegetation type: 1 – tundra vegetation; 2 – forest-tundra vegetation; 3 – forest vegetation (northern taiga, with propagation of permafrost soils); 4 – forest vegetation (middle taiga); 5 – forest vegetation (southern taiga); 6 – forest-steppe vegetation; 7 – steppe vegetation; 8 – desert and desert-steppe vegetation;

z2 – terrain relief: 1 – flat with a relative elevation of the relief (RER) up to 25 m; 2 – hilly with RER from 25 m up to 200 m; 3 – mountainous (low mountains area) with RER from 200 m up to 500 m, and with a prevailing slope gradient (PSG) from 5° up to 10°; 4 – mountainous (mid-mountain terrain) with RER from 500 m up to 1000 m, PSG from 10° up to 25°, and elevation above sea level of about 1000-2000 m; 5 – mountainous (highland area) with RER from 1000 m, PSG more than 25 °C, and elevation above sea level more than 2000 meters;

$z3$ – calculated soil moisture (CSM) – the ratio of the natural soil moisture (obtained in the calculated – spring period of the year) to the moisture at the liquid limit: 1 – low soil moisture with CSM up to 0.4; 2 – normal soil moisture with CSM from 0.4 up to 0.6; 3 – increased soil moisture with CSM from 0.6 up to 0.8; 4 – waterlogged soil with CSM from 0.8 up to 1;

$z4$ – content of sandy particles (expressed as percentage): 1 – sandy loam [50, 100); 2 – silty loam [0; 50); 3 – light loam, sandy [40; 100); 4 – light loam, silty [0; 40); 5 – heavy loam, sandy [40; 100); 6 – heavy loam, silty [0; 40); 7 – light clay, sandy [40; 100); 8 – light clay, silty [0; 40); 9 – heavy clay;

$z5$ – evaporation from the surface of the roadbed during summer months: 1 – extremely low, from 100 mm up to 150 mm (arctic deserts); 2 – very low evaporation, from 150 mm up to 200 mm (Siberian tundra provinces); 3 – low evaporation, from 200 mm up to 400 mm; 4 – average evaporation, from 400 mm up to 600 mm (taiga, Central and Central Black Earth regions of Russia, Krasnodar region); 5 – increased evaporation, from 600 mm up to 700 mm (mixed forests); 6 – high evaporation, from 700 mm up to 800 mm; 7 – very high evaporation, from 800 mm up to 900 mm (steppes); 8 – extremely high evaporation, from 900 mm up to 1000 mm (semi-deserts and deserts);

$z6$ – kind (type and subtype) of soil: 1 – light loam, large; 2 – light loam; 3 – silty loam; 4 – silty loam, heavy; 5 – light loam; 6 – light loam, silty; 7 – heavy loam; 8 – heavy loam, silty; 9 – sandy clay; 10 – silty clay; 11 – fat clay.

The following are integers encoding group characteristic features, associated to columns of the description matrix:

$z7$ – Selyaninov hydrothermal coefficient (SHC): 1 – excess moisture of soils with SHC from 1.4 up to 5; 2 – significant soil moisture in respective years with SHC from 1 up to 1.4; 3 – lack of soil moisture with SHC from 0.5 up to 1; 4 – arid areas with SHC up to 0.5;

$z8$ – plasticity number (PN, expressed as percentage): 1 – incoherent soils (sand, etc.) with PN from 0 up to 1; 2 – loam with PN from 1 up to 7; 3 – light loam with PN from 7 up to 12; 4 – heavy loam with PN from 12 up to 17; 5 – light clay with PN from 17 up to 27; 6 – heavy clay with PN from 27;

$z9$ – soil particle size (SPS, in mm): 1 – clay fraction with SPS up to 0.005; 2 – silty fraction with SPS from 0.005 up to 0.05; 3 – sandy fraction with SPS from 0.05 up to 0.5; 4 – large-sized fraction (coarse sand, gravel, boulders) with SPS from 0.5 up to 250;

$z10$ – average annual air temperature (AAT, in Celsius): 1 – extremely low temperature with AAT from -15.5 up to -10; 2 – very low temperature with AAT from -10 up to -6; 3 – low temperature with AAT from -6 up to -2; 4 – average temperature with AAT from -2 up to 2; 5 – high temperature with AAT from 2 up to 6; 6 – very high temperature with AAT from 6 up to 10; 7 – extremely high temperature with AAT from 10 up to 14.2;

$z11$ – snow cover depth (SCD, in cm): 1 – snowless areas with SCD up to 30; 2 – little-snow areas with SCD from 30 up to 50; 3 – moderate-snow areas with SCD from 50 up to 70; 4 – snow areas with SCD from 70 up to 100; 5 – extremely snowy regions with SCD from 100 up to 290;

$z12$ – soil freezing depth (SFD, in cm): 1 – shallow depth with SFD from 50 up to 180; 2 –average depth with SFD from 180 up to 220; 3 – great depth with SFD from 220 up to 260; 4 – very great depth with SFD from 260 up to 300; 5 – extremely great depth with SFD from 300 up to 600;

$z13$ – relative frost heave (RFH, %): 1 – non-heave soils with RFH less than 1; 2 – weakly-heave soils with RFH from 1 up to 4; 3 – heaving soils with RFH from 4 up to 7; 4 – strongly heaving soils with RFH from 7 up to 10; 5 – excessively heaving soils with RFH more than 10;

The following are integers encoding classification features, associated to columns of the distinctions matrix:

$k1$ – zone (1 – zone I, 2 – zone II, 3 – zone III, 4 – zone IV, 5 – zone V);

$k2$ – subzone (1 – subzone F, 2 – subzone H, 3 – subzone M);

$k3$ – road district (1 – district 1, 2 – district 2, 3 – district 3, 4 – district 4, 5 – district 5).

Frameworks of the article do not allow to present a representative fragment of description and distinctions matrices ($\mathbf{Q}^e$ and $\mathbf{R}^e$).

5. Data and knowledge base

The basis for the construction of the data and knowledge base for IIT RCZT is the above-described expanded matrix description of data and knowledge, as well as the description matrix of investigated territories. The data and knowledge base for IIT RCZT, which contains all characteristic features and their values, both in regards to the investigated territories and territories under investigation, as well as classification features of the investigated territories, is formed taking into account the peculiarities of natural-and-climatic conditions and the refined regularities of the influence of a geographic complex on the flow of the water-and-thermal regime of road structures.

To create IIT RCZT, the intelligent instrumental tool IMSLOG (previously created in the laboratory of intelligent systems of TSUAB [10]) is being further advanced. IMSLOG is designed to reveal various types of regularities, including fault-tolerant diagnostic tests; decision-making and justification with the use of cognitive tools.

The development of IIT IMSLOG is carried taking into account the specificity of the road-climatic zoning of territories.

In difference of the previously used concept "knowledge base - scheme of investigation – knowledge", the extended version of IIS IMSLOG uses the concept "project – scheme of investigation – data and knowledge base – knowledge". The modified subsystem of investigation templates is an interactive scheme of the investigation that includes: control elements (buttons, drop-down lists, text fields, etc.); display elements of graphic and text information (schedules, information windows, dynamic animation, cognitive graphics tools, etc..); report templates; summary, representing tables displaying the current status of calculations according to a presigned scheme of investigation; subsystem of security objects; file storage subsystem. Elements of the investigation template subsystem play a module of plug-ins in the previous version of IIT IMSLOG.

Methods of linear, nonlinear, and object-oriented programming, as well as technologies OLEDB and COM, have been used in the development of software modules. Software modules of the investigation scheme elements are stored in the library of algorithms, and are made in the form of plug-in dynamic modules. Each implements a given set of algorithmic and interface methods or methods of operation with data and knowledge. In difference from the unnamed input and output parameters of template elements used in previous versions, in the advanced version IIS IMSLOG each element of the scheme has a set of named parameters used to exchange data between modules. Software modules have been developed in C ++ in the environment Embarcadero RAD Studio XE.

6. Conclusion

It is proposed to construct a database and knowledgebase for the intelligent information technology of road-climatic zoning of territories on the basis of the proposed extended matrix description of data and knowledge, as well as on the basis of the description matrix of territories under investigation.

The structurization of data and knowledge on the road-climatic zoning of territories, necessary for the development of IIT RCZT, based on test methods of pattern recognition and cognitive graphics tools is given in the paper. The structurization was fulfilled with the use of reference data as well as data collected in the course of long-term field and laboratory studies in the territory of Western Siberia [3 – 5].

For the first time, intervals of values of characteristic features have been separated and their integer values used in the description matrices have been given.

Formation of the data and knowledge base on the base of the new concept "project – investigation scheme – data and knowledge base – knowledge" proposed in the developed IIT IMSLOG that will be used to construct of the intelligent system RCZT, which is the base of creating IIT DKRT, has been given.

The developed IIT IMSLOG can be used in the construction of intelligent systems to solve a wide range of diagnostic, prognostic, and organizational-administrative tasks from various problem and interdisciplinary areas. Reasonability of its use for the diagnostics of the stability of structural materials, including complex composite materials, to various types of deforming effects is beyond

question. For these purposes, first of all, it is necessary to carry out structurization of data and knowledge in regards to deforming effects according to the matrix model of data and knowledge representation, used in intelligent systems based on test methods of pattern recognition, support of the diagnostic decision-making, and tools of cognitive graphics for justification of decision making results [9]. Further, it is necessary to create a data and knowledge base, as well as to construct an application intelligent system on deforming effects on the basis of IIT IMSLOG.

References
[1]   The national program of modernization and development of automobile roads of the Russian Federation until 2025. Project from March 25, 2004 / Ministry of Transport and Communications of the Russian Federation. – Moscow, 2004.
[2]   34.13330.2012. Set of rules. Automobile roads. The updated edition of SNIP 02.05.02 – 85 *. Automobile roads. Gosstroy of the USSR. – Moscow: CISD of Gosstroy, 2004 (approved by the Order of the Ministry of Regional Development of Russia, dated 30.06.2012, No. 266).
[3]   Efimenko V. N., Efimenko S. V., Sukhorukov A. V. Technology for Assignment of Predicted Behavior of Subgrade Soils in the Design of Road Pavements. Key Engineering Materials. Vol. 683 (2016). pp. 250-255.
[4]   Efimenko S., Efimenko V., Sukhorukov A. Peculiarities of strength and deformability properties of clay soils in districts of Western Siberia. ADVANCED MATERIALS IN TECHNOLOGY AND CONSTRUCTION (AMTC-2015): Proceedings of the II All-Russian Scientific Conference of Young Scientists «Advanced Materials in Technology and Construction». Vol. 1698 (2016). 070020.
[5]   Efimenko V. N., Efimenko S. V., Sukhorukov A. V. Accounting for natural-climatic conditions in the design of roads in western Siberia // Sciences in Cold and Arid Regions. Vol. 7. Issue 4. August. 2015. pp. 307-315.
[6]   Russam K., Coleman J.D., The Effect of Climatic Factors on Subgrade Moisture Conditions // Geotechnique, XI, 1961, № 1, P. 22–28.
[7]   Zapata C.E., Houston W.N., Calibration and validation of the enhanced integrated climatic model for pavement design. – Washington, D.C.: Transportation Research Board, 2008. – 62 p.
[8]   Yankovskaya A., Efimenko S., Cherepanov D. Structurization of data and knowledge for the information technology of road-climatic zoning // Applied Mechanics and Materials. Vol. 682. 2014. – P. 561-568.
[9]   Yankovskaya A.E. Logic tests and tools of cognitive graphics. Publishing house: LAP LAMBERT Academic Publishing, 2011. – 92 p.Yankovskaya A.E., Gedike A.I.,
[10] Ametov R.V., Bleikher A.M. IMSLOG-2002 Software Tool for Supporting Information Technologies of Test Pattern Recognition // Pattern Recognition and Image Analysis. Vol. 13. 2003, № 4. – P. 650-657.