Information System to Analyze Human Environmental Wellbeing for Assessing and Reducing Technogenic Risk

O E Bezborodova¹, O N Bodin² and B V Chuvykin¹

¹Penza State University, 40, Krasnaya Street, Penza, 440026, Russian Federation
²Penza State Technological University, 1a / 11, Pr. Baidukova / St. Gagarina, Penza, 440039, Russian Federation

E-mail: oxana243@yandex.ru

Abstract. The article provides a rationale for the need of developing the information system to analyze human environmental wellbeing (HEW) on the basis of artificial neural networks. The application of artificial neural networks is pre-conditioned by the need of prompt analysis and processing to take decisions on the implementation of environment protection measures for large data arrays accumulated as a result of environmental monitoring. The suggested information system for technogenic risk assessment implements a multi-limit way to control the state of a multi-parametric object and allows assessing human environmental wellbeing in real time. The authors provide structural charts for the information system to assess technogenic risk and the block to process and store information on the basis of an artificial neural network LVQ (Learning Vector Quantization). The authors provide an algorithm and the results for the artificial neural network LVQ training.

1. Introduction

According to the data of World Health Organization (WHO) [1] and taking into account the concept “environment quality” given in [2], the sources of risks to human health are divided into four groups. The “impact” of the environment state in the place of residence is assessed by WHO as twenty five percent; it consists of chemical, physical and biological factors (see figure 1). Figure 2 provides statistic data [3] on air pollution in Russian cities.

Figure 1. Sources of risks to human health according to the WHO data.

Figure 2. Exceeding the limit concentrations of chemical substances in the air of Russian cities, times [3].
The analysis of medical statistics data for the last decades [4] shows that the impact of chemical substances results in the increase in the incidence of chronic respiratory diseases, hepatic disorders, kidney diseases, toxic and allergic processes, genetic defects and congenital malformations, malignant tumors and hemopathy, vegetovascular disorders, neuroses, heart and arterial sclerotic diseases, i.e., the dysfunctions of all human organs and systems.

Today there is a great number of systems used to control the environment state. Mostly, they control climatic parameters (environment temperature, humidity and air velocity) [5-8], but only some of them control the content of chemical substances [9-11]. Comparative analysis for these systems is provided in Table 1.

| Comparison parameters | Controlled parameters | Composite indicator use | Opportunity | Structural solutions |
|-----------------------|-----------------------|-------------------------|-------------|---------------------|
| System name           | chemical climatic     | remote control          | measureme nt result processing |                     |
| SIPA-N [5]            | A, H                  | –                       | –           | SP                 |
| ASMC [6, 7]           | S                     | –                       | –           | SP + MP            |
| IVSMPSA [8]           | A                     | –                       | –           | MP                 |
| Automatic air         |                        |                         |             |                     |
| monitoring system     | “Atmosfera-S17” [9]   | CO, N₂O, NO₂, NOₓ, SO₂, H₂S, C₅H₆, TSP | +           | SP + MP            |
|                       |                       | CO, NO, NO₂, NOₓ, NH₃, THC, CH₄, C₂H₆, C₄H₁₀, C₆H₁₄, C₇H₈, C₈H₁₀, CO, NO-NO₂-NOₓ, NH₃, SO₂, H₂S, C₅H₆, C₇H₈, CH₄ | – | – | SP |
| SLAV [10]             | A                     |                         | +           |                     |
| ASAM [11]             | A                     |                         |             |                     |

A – atmosphere, H – hydrosphere, S – soil, SP – stationery post, MP – mobile post.

The analysis of functional capabilities of the existing systems for the control of the environment states shows that they are measurement systems only and cannot provide enough data on the impact technogenic objects have on the environment nor they process data for taking prompt control decisions to ensure human environmental wellbeing (HEW).

The article purpose is to develop an intellectual information&measurement system to assess HEW on the basis of the technogenic risk analysis by means of data clustering to reveal the discrepancies between the actual values of atmospheric and hydrosphere parameters and the limit values as well as to form the list of environment protection measures.

2. Problem statement

The environment quality control is under a close attention in the Russian Federation. Since January 1, 2019 the provisions of the law [2] came into force obliging technogenic objects to equip all the stationery sources of emissions and pollution discharge with the means for the automatic measurement and account of pollution volumes as well as by technical attachment means, equipment for data transmission on the values of emissions and pollution discharge to the state register of objects having negative impact on the environment. This provides the basis for forming large arrays of measurement results which should be processed in real time with a high veracity level. The analysis and prompt data
processing capabilities are connected with the application of new technologies “Internet of things” (IoT), “Big Data” (BD), “Cloud Computing” (CC) and artificial neural networks (ANN).

BD include the technologies for processing structured and non-structured data with a continuously increasing volume. The key characteristics of BD are the volume, variability, speed and value. The data variability allows identifying the dependencies in those areas where they cannot be seen otherwise [12, 13].

The technology of BD processing, obtained in the course of environmental monitoring, should include the following stages. Most importantly, the data should be collected in network attached storages which can be both isolated and united into a single system. The data should be duplicated in a mandatory way to eliminate the risk of possible losses. The data accumulated in such a way are characterized by the structure absence, i.e., this can include text, images in various formats, etc. Further, one needs to conduct data machine processing when no human involved in the process. Such data processing stipulates for the use of the ANN after deep learning which model high-level abstractions into the data using architectures consisting of a set of no-linear transformations. The speed of data processing in the ANN is compared with the real time pace. This provides an opportunity to work not with static data but with a continuously incoming data flow. Such data processing consists in their structuring and comparison with pre-set limit values (LV). By the comparison results, the decision taking person is offered a set of environment protection measures the implementation of which will improve HEW.

At present RF employs a system for environment quality and population health assessment based on the principles of compliance with the set sanitary and hygienic norms (limit values, or LV) [14, 15, 16]. However, it does not allow for considering hidden health impacts because of chemical substances. Today this harm cannot be assessed in the generally comprehensible quantitative environmental & economic parameters. This is a significant drawback of the existing rationing system.

To eliminate this drawback, it is necessary to develop and integrate the methodology for assessing potential harm to population health - a procedure based on the quantitative assessment of technogenic risk - into the system for analyzing the safety of technogenic object activity. In its turn, it brings about the need to develop the methods for the application of quantitative assessment and forecast of the technogenic risks for HEW.

That is why it is necessary to develop and test the methods for the technogenic risk assessment and forecast with the ANN application. Such methods will allow forecasting with the account of the territorial technosphere peculiarities regardless of the criteria selected for HEW assessment. On the basis of the forecast one can form a list of environment protection measures aimed at the reduction of technogenic risks.

3. Theoretical part
The authors suggested the HEW concept based on data mining [17] which has no above-mentioned flaws and provides a proper theoretical rationale for building the information system to assess technogenic risk.

The information system of HEW analysis, used to assess and reduce technogenic risks, is built on the ANN basis. The theoretical rationale for the neural network approach to the generation of the HEW analysis information system is provided by the derivation from the Kolmogorov–Arnold representation theorem which was proved by Hecht-Nielsen. According to him, the derived continuous function of some variables can be approximated by a neural network with any preset accuracy rate [18]. The key ANN component is the Kohonen layer. It consists of adaptive linear combiners. The output signals of the Kohonen layer are processed with the account of input adder weights by the rule “the largest impulse is turned into a unit impulse, the others are turned to zero” [19].

The ANN offered implements the method suggested by authors in [20]. To assess HEW, this method considers the human environment as a multi-parametric object characterized by a large number of parameters the values of which change under the impact of technogenic objects. To control every parameter, one should set several limit values and assess each parameter in relation to them.
Further, one defines a composite value for the multi-parametric object state that shows its current state (normal, permitted, critical, collapse). The ANN analyzes the measured or calculated values by the decision procedures. Relating the measured (calculated) values to any of the ranges is made by the following rules:

– setting a necessary number of limit values, for example, on the levels of 0.5 LV and 0.7 LV;
– determine the conformity of the values of the signs of compliance of the measurement results and permitted, limit and out-of-limit signs $\delta_{ij}$ to the set ranges of permitted, limit and out-of-limit signs by the formula:

$$\delta_{ij} = \begin{cases} 0, & y_{nij} \leq y_j \leq y_{wij} \\ 1, & y_j < y_{nij+1} \\ \frac{y_j-y_{nij}}{y_{nij+1}-y_{nij}}, & y_{nij} < y_j \leq y_{nij+1} \\ \frac{y_j-y_{nij}}{y_{nij+1}-y_{nij}}, & y_{nij+1} \leq y_j < y_{nij} \end{cases}$$

where $y_{nij}$, $y_{wij}$ – lower and upper limits for each control value range;
– defining composite indicators of the multi-parametric object state $\Delta_i$ for each set range of values by the formula:

$$\Delta_i = \sqrt{\sum_{j=1}^{n} \delta_{ij}^2}$$

The result of the neural network analysis for the multi-parametric object state is a composite indicator of the multi-parametric object state.

Depending on the selected assessment scale the system forms a list of environment protection measures the implementation of which reduces technogenic risks.

### 4. Practical implementation

To define and assess human environmental wellbeing on the basis of neural network analysis, the authors suggest using the information&measurement system, the structural chart of which is provided in figure 3. The structural chart of the block for information processing and storage on the basis of ANN $LVQ$ (Learning Vector Quantization), implementing the above-mentioned method, is given in figure 4.

The measured and processed values of the controlled parameters of the multi-parametric object are sent to the ANN input.

The signal value on the input of ANN $LVQ$ is defined by the formula:
\[ \Delta_n = F_{lin} \left( \sum_{j=1}^{S} w_{jk}^2 \cdot F_{comp} \left( \sum_{i=1}^{N} (y_j^c - w_{im}^1)^2 \right) \right) \]  

(3)

where \( y_j^c \) – \( j \)-th component of the input vector; \( w_{im}^1 \) – \( i \)-th component of the weight vector of the \( m \)-th neuron of a competitive layer; \( w_{jk}^2 \) – \( j \)-th component of the weight vector of the \( k \)-th neuron of a linear layer; \( F_{comp} \) – transfer function of a competitive layer identifying the winning neuron; \( F_{lin} \) – linear function of the activation of the distributing layer neurons; \( N \) – dimensionality of the input vector of the neural network; \( S \) – neuron number in the competitive layer; \( \Delta_n \) – value of the \( k \)-th output of the neural network.

The method of the neural network analysis of the technogenic object parameters to define HEW uses the \( j \)-th ANN LVQ (one ANN for each controlled parameter).

For a successful ANN functioning, a training algorithm is necessary. The suggested algorithm is provided in figure 5.

The algorithm in Figure 5 uses the following designations:

where \( n \) – number of inputs of the neural network; \( t \) – No. of the epochwise training; \( \beta(t) \) – monotonically increasing function changing from 0 to 1 in the course of training; \( x' \) – modifying training vector, \( d_1 \) – distance from \( x'h \) to the 1st winning neuron; \( d_2 \) – distance from \( x'h \) to the 2nd winning neuron; \( \varepsilon \) – width of the window where a training vector should fall \( x_h \), the value \( \varepsilon \) is selected from the range 0.2 ÷ 0.3. \( m \) – winning neuron number; \( \delta \) – parameter impacting the degree of correction / weight repulsion, is in the range 0.1 to 0.3; \( \eta \) – parameter impacting the speed of adjusting the neuron weight in the competitive layer. The algorithm stipulates that the \( LVQ \) value \( \eta \) is changed in a mandatory way from 1 to 0 during training. In the course of training \( \eta \) is assumed to be 1.

To define the quality values of ANN training, the authors conducted the research in the course of which ANN should be trained multiple times under various training parameters. After each training procedure the authors calculated the specificity (type 2 error), sensitivity (type 1 error), generalization error, sampling error, training error, training time. The examples of neural network training: are provided in figure 6.
5. Conclusions
The ANN trained in such a way promptly analyzes large data arrays dividing them by the set criteria into clusters corresponding to a certain HEW level (normal, permitted, critical, collapse). In case of revealing the fact of the discrepancy with the standard value (critical, collapse) the system suggests a list of environment protection measures improving HEW in the area of technogenic factors impact.

6. References
[1] Stevens G, Mascarenas M, Mathers K 2009 Global health risk factors: progress and challenges Bulletin of the World Health Organization Rel. 89 9 pp 645-732 https://www.who.int/bulletin/volumes/87/9/ru/
[2] Federal Law "On Environmental Protection" dated 10.01.2002 № 7-FL (with changes and additions from 30.12.2020 № 494-FL)
[3] Federal Service for Hydrometeorology and Environmental Monitoring (Roshydromet). Official site About emergency, extremely high and high environmental pollution on the territory of Russia http://www.meteorf.ru/product/informmaterials/99/?year=2020&ID=99
[4] Gichev Yu P 1995 Human health as an indicator of the environmental risk of industrial regions Bulletin of the Russian Academy of Medical Sciences 8 pp 52-54
[5] System for measuring atmospheric parameters SIPA-N Official site https://elins.ru/production/nodes/sistema-izmereniya-parametrov-atmosfery-sipa-n/
[6] Kiselev M V et al 2018 IOP Conf. Ser.: Earth Environ. Sci. 190 012031
[7] Bazarov A et al 2018 Mobile Measurement System for the Coupled Monitoring of Atmospheric and Soil Parameters Russian Meteorology and Hydrology 43 271-275 10.3103/S106837391804009X
[8] Korolkov V A et al Pilot project of measuring and computing system for mesoscale monitoring of atmospheric boundary layer 21st International Symposium on Atmospheric and Ocean Optics: Atmospheric Physics, edited by G G Matvienko, O A Romanovskii Proc. of SPIE Vol 9680, 96805Z © 2015 SPIE CCC code: 0277-786X/15/$18 do: 10.1117/12.2205475
[9] Hardware and software complex "Atmosphere-C17" Official site. http://atmosphere.trans-e.ru/
[10] Automatic air monitoring stations Official site http://www.nevaline.com.ru/resheniya/avtomaticheskie-stantsii-monitoringa-atmosfernogo-vozdukh.html
[11] Automatic atmospheric monitoring station Official site. https://www.saveplanet.su/tehno_338.html
[12] Duke V A Data Mining Information technology: website URL: http://www.inftech.webservis.ru/it/database/datamining/ar2.html
[13] Khaikin S 2006 Neural Networks: A Comprehensive Foundation (M.: «Williams») p 1104
[14] R 2.1.10.1920-04 Guidelines for Assessing Public Health Risks from Exposure to Environmental Polluting Chemicals

[15] Kovalevskaya O Yu, Blinovskaya Ya Yu, Agoshkov A I, Vasyanovich Yu A, Petukhov V I, Doryshev Yu S 2013 The risk of emergencies during the operation of offshore oil platforms Problems of the development of geo-resources of the Far East Issue 4: Mining information and analytical bulletin (scientific and technical journal) Selected articles (special issue) 12 p 144 (M.: publishing house "Gornaya Kniga") pp 3 – 11

[16] Zhigula L D, Petukhov V I, Zhigula E A 2013 Criteria for identifying natural hazard factors and problems of risk management in the field of recreation Modern problems of science and education 3 URL: http://science-education.ru/ru/article/view?id=9182

[17] Bezborodova O E et al 2021 IOP Conf. Ser.: Earth Environ. Sci. 666 022004

[18] Hecht-Nielsen R 1987 Kolmogorov’s mapping neural network existence theorem IEEE First Annual International Conference on Neural Networks (San Diego) Vol 3 pp 11-13

[19] Kohonen T 1989/1997/2001 Self-Organizing Maps (Berlin – New York: Springer-Verlag) First edition 1989, second edition 1997, third extended edition 2001

[20] Method and system for complex monitoring of the state of a multiparameter object based on heterogeneous information: pat. 2719467 declared 11.11.2019 published 17.04.2020 Bul. 11 44 p