Using Tools of Intellectual Analysis In Area Safety Management

Evgeniy Materov[1] and Valery Nicheporchuk[2]

[1] *Siberian Fire and Rescue Academy of EMERCOM, 1, Severnaya, Zheleznogorsk, 662972, Russia*
[2]*Institute of Computational Modelling of the Siberian Branch of the Russian Academy of Sciences, 50/44, Academgorodok, Krasnoyarsk, 660036, Russia*

Abstract

In this research, we are considered the different methods of analytical data processing used for support information tasks for managing the safety of territories. The problem of data availability with use data lakes technology for operating analytics was investigated. The use of Data Mining, Machine Learning to jointly processing Big Data of monitoring, hazard event catalogs and territories characteristic allows us to fend off threats quickly and effectively to people life. The results of spatial modeling of fire risks are described. A method for predicting time series using artificial neural networks is proposed.

Keywords

Data analysis, preprocessing, spatial visualization

1. Introduction

The global informatization and digitalization of human activity has simplified the collection and production of information. However, it is still not possible to achieve the desired improvement in the management quality, decrease in the number of personnel and specialists in decision making. The problem of developing techniques for applying modern information technologies in different practical activities is rather urgent. One of the complex spheres which requires the deep analysis of big data volumes, application of different analytical and situation models is the management of natural and technogenic safety of areas [1, 2]. Unlike other spheres of management, decisions concerning responses to situations which seldom occur in everyday life, are made under the conditions of time deficiency. A great number of factors have to be taken into account, as well as their interaction, probability of occurrence, extent of influence on the outcome of the situation or their influence on the overall safety conditions. Here, the price of wrong decisions is extremely high.

The present study is a brief review of the techniques and services of intellectual data analysis (IDA). It presents the investigation results of various data sets used in the processes of information support of the management of natural and technogenic safety of the region and municipal districts. Tools of analytical modeling with different degrees of complexity have been suggested which allow one to accelerate implementing data-based management.

The paper is organized as follows. Section 2 we review of some manage tasks of providing natural and technogenic safety. Section 3 presents a description of some examples of analytical modeling. Finally, we conclude the paper in Section 4.

2. Intellectual analysis in the tasks of providing natural and technogenic safety

In spite of a great variety of definitions (data mining, in-depth analysis, etc.) almost all the IDA methods are aimed at finding information in the data the which is necessary for decision making, this

SibDATA 2021: The 2nd Siberian Scientific Workshop on Data Analysis Technologies with Applications 2021, June 25, 2021, Krasnoyarsk, Russia
EMAIL: materov@gmail.com (E. Materov); valera@icm.krasn.ru (V. Nicheporchuk)
ORCID: 0000-0002-3916-0613 (E. Materov); 0000-0001-5365-1307 (V. Nicheporchuk)

knowledge being non-trivial, practically useful, and easy to interpret. The use of analysis for big data is based on the integration of mathematical tools and modern information technologies.

N. Chen et al. [3] analyzed more than 170 publications describing various applications of computational intelligence technologies in the emergency management. The above mentioned study shows the application of decision trees, artificial neural networks, support vectors, evolutionary algorithms and other technologies in particular tasks of emergency management. In [4], emergency management is described as a '4R' process, namely, reduction, readiness, response and recovery, where reduction is referred to pre-incident phase, readiness and response are referred to during incident phase, and recovery is referred to post-incident phase. In each phase, the outcome of decision-making impacts significantly the evolution of incidents and the effectiveness of emergency management.

Many authors concentrate their attention on the analysis of messages and news feed of social networks (Twitter, Facebook) as a source big data [5, 6]. Despite the peculiarity of the processing tools, social networks can be considered as a source of monitoring data, instantly responding to technogenic and natural emergencies. The role of this source of data in decision making at the federal and regional level has considerably increased in the period of the COVID-19 pandemic. Here, one should take into account the competence of the authors of messages who make hasty conclusions based on incomplete and inaccurate information and with the lack of special knowledge.

A considerable amount of investigations is being conducted by the Russian Ministry of Emergency Situations (EMERCOM). The problems of computational modeling, forecasting emergencies, ways of increasing the efficiency of response to accidents and incidents are under study. Along with the increasing area of the real-time monitoring there is a task of applying promising methods of big data analysis for the management of risks. A technological basis has been established in the form of data lakes implementing interagency information exchange. Investigations are being carried out concerning the techniques of complex processing of information about events, objects and processes; these techniques allow one to make efficient decisions to prevent emergencies, to decrease the risk of fires, technological accidents and natural disasters.

The main source of big data in the field under consideration is the complex monitoring of natural and technogenic emergencies and hazards. Data processing is used for forecasts of different urgency degree, for early prevention of emergency situations, and for providing readiness for prompt response in the case of emergency. Real-time complex processing of the data concerning daily monitoring of environment across vast land areas, results of controlling a great number of industrial facilities in combination with the characteristics of the facilities and processes allows one to considerably improve the quality of safety management of the areas. [7]. The application of different types of analytical modeling is shown in the Table 1.

Table 1

Methods of analytical solution of the problems connected with providing regional safety

Initial data	Type of analysis	Visualization	Management
Monitoring of conditions	Revealing anomalies	Dynamic maps of the area, cross-tables, graphs	Early warning about hazards and threats
	Predictive	The same	Transfer the safety system to the high alert mode
Catalogue of events	Clusterization of periods	Diagrams, heat maps	Preparation for cyclic emergencies
	Clusterization of areas	Cartograms of hazards for the whole area	Substantiation of strategic preventive measures
	Clusterization according to the scale	Dynamic maps, cross-tables	Introducing changes in the operational plans of actions and in the protection system
Single event	Analysis of the	Gantt charts,	Acquisition of knowledge.

Initial data	Type of analysis	Visualization	Management
	sequence of actions	business-processes	Updating the operational plans of actions
	Analysis of the occurrence of dangerous factors	Large-scale dynamic maps, cross-tables, diagrams	The same
Monitoring and events	Cause-consequence	Graphs, maps	Substantiation of the techniques of minimizing hazards

The improvement of the safety management decisions occurs due to synergy effect: the volume of information resources is increasing, technologies of processing, modeling and presenting data are being improved. Devices for collection of information become more widely spread which is accompanied by the technological revolution in data storage. In the 1990-ies, databases were replaced by specialized data storages oriented to analytical tasks [8]. At present, Hadoop technologies of data lakes etc. are being actively developed. The possibility of dealing with various types of "raw data" allowed solving the problem of interagency information exchange which for a long time had complicated the processes of digitalization in the regional safety management. It became easier to organize complex on-line monitoring. Data get into the lakes by means of uploading them in intervals or in a continuous stream. The low cost of the technology allows scaling data lakes both according to the issue of interest and according to the region (e.g., federal districts of Russia). This technique is already used in the open joint-stock company "Russian Railways" [9].

Data lakes provide opportunities for collecting and storing any data, including raw unprocessed data for any period of time and processed/purified data, for making deep analysis and considering the data, providing the flexible access to various data and scenarios. The advantages of using data lakes over data storages depend on a particular task. In organizing complex real-time monitoring of emergencies and inter-agency exchange at the federal level, it is preferable to use the new technology. In this case, each agency collects observation data in the frames of the branch (corporative) monitoring system and transfers these data as raw or minimally processed information. The recipient can use some part of information which considerably saves the processing time and decreases the significance of failures in the case of changing data formats. At the regional level, with the rarely changed structure of information streams and full space-time coverage of the monitoring it is worth using data storages.

In the National Center for Crisis Management (NTsUKS), data lakes have been used since 2020. There exists the access to real-time and archive data of environmental monitoring, and data concerning hazardous facilities as well as access to the reports of Federal Agency for hydrometeorology, Rosatom, Federal Agency for Forestry, Ministry of Internal Affairs and other agencies. It is urgent to develop new techniques of data analysis and visualization taking into account the requirements of a particular person who makes decisions (PMD) in the hierarchy of the regional safety management. In providing the guaranteed access to regularly updated data it is necessary to implement the technique of transforming data into knowledge. It allows providing information of the necessary volume, content and type for every PMD. As is seen in the practice of risk management and response to emergencies and accidents, using the traditional information systems did not prevent PMD from having incomplete and inaccurate information, but, in addition, it created the problem of excessive data, which required time and experience in order to comprehend these data under the conditions of stringent requirements to the decisions being made.

It is rather difficult to distinguish between intellectual and "usual" data analysis. Modern technologies allow one to quickly process large arrays of data and the results of the analysis aid in making well-grounded and balanced decisions. In [10] the authors give the characteristics of the descriptive, exploratory, inductive, prognostic and cause-consequence types of analysis as well as the peculiarities of the detailed process modeling. The most complex ones and the most desirable in the area management are the last three types of the analysis. The application of the methods of cause-consequence analysis to estimate the factors of different natural and technogenic hazards is considered in [11]. This study gives examples of the predictive analysis of the statistics of dangerous events accumulated in NTsUKS of EMERCOM, Russia.

In the processes of big data analysis, ready software environments are used (desktop applications and web-services) as well as programming languages which allow developing problem-oriented products. Each approach has its advantages and drawbacks. Today, the human factor (qualification of users, designers and developers and readiness for digitalization of business processes) are often more significant than financial investments into the digital infrastructure.

The most popular analytical platform for business analysis is MS PowerBI. The interfaces of connectors (gateways), desktop application and web-services are identical and user-friendly for those working with Microsoft products. The xVelocity module provides column-wise compression of data and in-memory computing; this approach provides better aggregate functions than OLTP-systems. The peculiarity of Power BI is an open interface for connecting various visualizations including side developers.

Tableau is a software for data visualization and business decision making, which is the leader in the segment of the BI platform of 2012-2019. It provides the possibility to collect data from relational databases, OLAP cubes, spreadsheets, cloud databases, including social networks and to create visual representations for these data. It is compatible with popular spatial data formats for visualizing spatial objects, creating dynamic maps and cartograms.

Programming languages have a higher entry threshold, but they have the necessary flexibility in solving applied problems of Data Science, machine learning, and in fulfilling other tasks. The leaders in the use of intellectual analysis are Python, R, and Julia. Python is the most popular, it has a wide range of applications, and it is easy to learn. The R language provides libraries allowing statistical processing and data visualization, creating interactive reports, dashboards, and it has a well-thought-out ecosystem. Julia is characterized by a high speed of operation. The languages are free and open source software environments [12].

3. Examples of solving analytical problems

The means of intellectual analysis in the field of natural and technogenic safety are used to solve the problems of identifying factors of occurrence and escalation of emergencies, forecasting the situation, searching and substantiating optimal methods of protection, etc. The use of IDA (intellectual data analysis) methods makes it possible to solve the problems of generalization, analytical processing, algorithmization and presentation of data in numerical and visual forms. For example, the mobile application “Thermal Points” of the information and analytical center of EMERCOM of Russia was developed using the methods of machine learning and big data analysis. Based on the analysis of temperature anomalies in satellite images, the probability of fires and the class of fire hazard of territories are estimated. The results are published on the web portal, mobile application and sent to the regions of Russia through available means of communication.

Among the tools of statistics, reporting, business analytics, machine learning in the application to data lakes at the initial stage of the research “Development of proposals for the use of promising methods of Big Data Analytics in the activities of the EMERCOM of Russia”, the main focus is on the Amundsen metadata service. It is used to describe the structure of databases, indexing information resources, including tables, dashboards, streams, etc. This increases the performance of analytics when interacting with data, and avoids the Data Swamp, which makes it difficult to work with haphazardly recorded data. The Apache Superset open source cloud application is used to explore and visualize data in the form of dashboards, namely “data marts”. The application is capable of handling data on a petabyte scale. Local samples for test modeling were saved in the free PostgreSQL DBMS.

The problem of preliminary processing of spatial characteristics of events and objects for cartographic visualization of results and performing spatial analysis is solved. The software used to collect information, as a rule, was developed without taking into account the capabilities of GIS. Spatial data on man-made, domestic and natural fires, and other types of emergencies are presented in the form of a building address, a road distance, a forest block number, etc. A reference address base and an address geocoding service have been developed. The Python libraries used are regex (string preprocessing), nltk (address comparisons) and fuzzymatcher (evaluation of the difference between the strings using the Levenshtein distance). To visualize events and the results of their processing, the capabilities of the R language are used such as the Leaflet library for interactive map development and ggplot2 library for static mapping based on the OpenStreetMap resource. Fire hazard maps of Siberian metropolitan cities have been created, and the areas with increased risk are highlighted by points and core densities. Filtering allows setting time periods and characteristics of the events (consequences, duration, resources involved in emergency management, etc.). The ability to test hypotheses related to geodata is provided. It concerns, for example, the dependence of the number and scale of fires on the density and type of the building system, the availability of centralized water supply systems, the distance to fire departments.

Figure 1 shows the “digital image” of the city of Krasnoyarsk. With the use of kernel densities, the dynamics of the development of fires has been analyzed, zones with decreasing and increasing fire hazard trends have been identified. The developed analytical tool is being implemented in the Federal State Fire Inspectorate to support preventive activities.

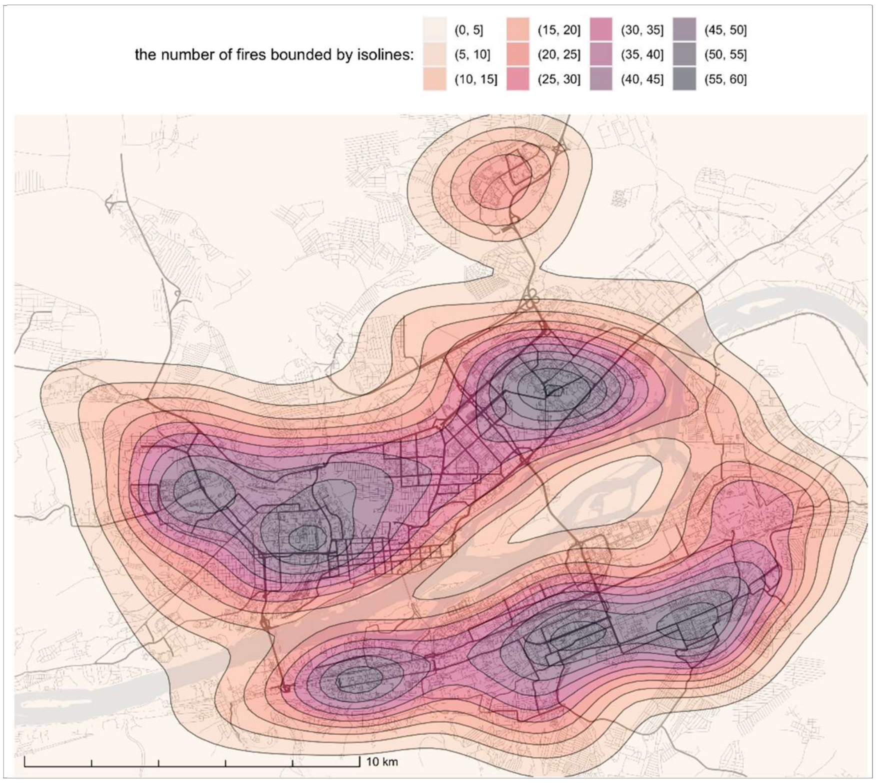

Figure 1: Density of fire events in the Krasnoyarsk city in 2019

Figure 2 shows an example of visualizing the transport accessibility of socially significant objects in response to fires and accidents. The map is built using the R language osmdata library, implemented based on the Overpass API. The standard time of arrival from the moment the fire brigade leaves the fire department to the burning facility in the city is 10 minutes. So far, the accessibility of the objects has been calculated without taking into account traffic and the condition of the road surface. The results of analytical processing are used to optimize the fire coverage and substantiate management and methodological decisions in the field of the area safety.

The time series of the number of dangerous events, such as the number of domestic fires and increased water levels, have been investigated. For short-term and medium-term forecasting, the R language modeltime library is used, which is based on the tidymodels machine learning framework. It allows to create forecast models across multiple frameworks, create tables of metrics, and to automatically select the best model for a specific set of data. The library supports time series ensembles, AutoML for H2O algorithm models, GluonTS-based deep learning, and more. Other R libraries allow one to perform STL time series decompositions for determining trends and seasonal components, and to consider time series anomalies.

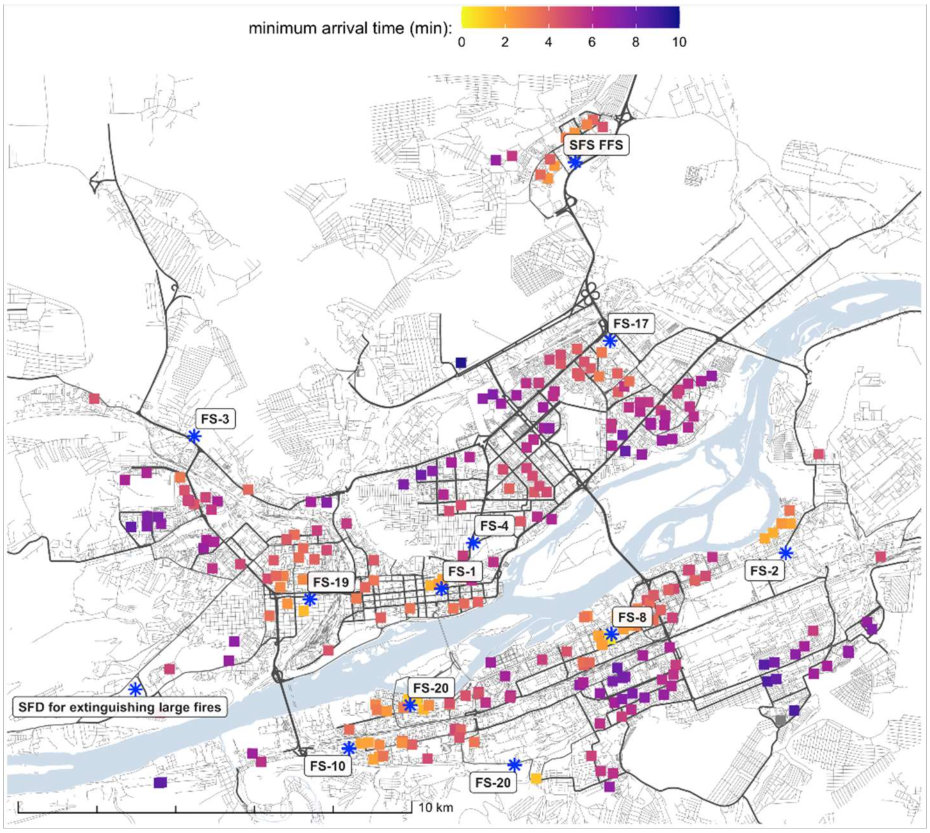

Figure 2: Accessibility of the kindergartens in Krasnoyarsk for fire and rescue units. (FS – Fire Station, SFS – Special Fire Station, SFD – Special Fire Department)

Figure 3 shows the hydrograph of the Yenisei-Yartsevo water gauge, as calculated by different models. To prepare for the spring flood, it is important to know the time of the onset of its peak, the scale of flooding and the duration of the water stand. In other words, forecasting accuracy is critical only for 1-2 weeks a year. Similar approaches are used for hydrological forecasting of low water in the Arctic zone. Knowledge of the river regimes is necessary for making decisions on the delivery of goods and fuel to remote regions of Siberia, estimating the required loading of ships and logistic schemes.

Figure 3: Comparison of the water levels estimated by different machine learning models

An important step in presenting the results of the described algorithms and visualizations is the report. In the conditions of excessive data, increased requirements are imposed for it to be visible and accessible for understanding by PMD. Static web page reports are developed in flexdashboard based on R Markdown while interactive web pages are based on Shiny. HTML widgets are used to build graphs, charts, and interactive tables. A web report on fires in the Krasnoyarsk Region is available at https://fire-report-2019.netlify.app/. It is planned to create more complex web services which allow using the developed methods while supporting decision-making. Examples including the complete code, models, and data links for computational reproducibility can be found in the author's blog at https://materov-blog.netlify.app/. It is shown that in the presence of sufficiently powerful tools in the form of Python and R programming languages, it is possible to implement data-based management in the field of the area safety.

4. Conclusion

Using analytical processing technologies in combination with services for the consolidation of complex monitoring data has increased the efficiency of territorial safety management. The tasks of a comprehensive description of the safety characteristics of territories; updating big data and monitoring their reliability; analytical processing by specialized algorithms; visual dynamic presentation of results; formalization of the experience of crisis management are set and partially solved.

The analytics used at all territorial administration levels will allow not only to reduce losses from emergencies through a deep justification of preventive measures, but also to transform change the structure of the territorial security system, implement data-driven governance.

5. References

[1] V. V. Moskvichev, I.V. Bychkov, V.P. Potapov, O.V. Taseyko, Yu. I. Shokin, Information system of territorial management of development risks and security, Bulletin of the Russian Academy of Sciences 87(8) (2017) 696–705. doi: 10.7868/ S0869587317080035.

[2] V. V. Nicheporchuk, A. I. Nozhenkov, The technology of situational modeling of dangerous events for territorial management information support, Procedia Structural Integrity 20 (2019) 248–253.

[3] N. Chen, W. Liu, R.Bai, et al., Application of computational intelligence technologies in emergency management: a literature review, ArtifIntell Rev. 52 (2019) 2131–2168. doi: 10.1007/s10462-017-9589-8.

[4] Y. Chen, Y. Lin, Controlling the movement of crowds in computer graphics by using the mechanism of particle swarm optimization, Appl Soft Comput. 9 (2009) 1170–1176.

[5] S. Mei, Emergency Management of Network Public Opinion Events based on FPGA and Machine Learning, Microprocessors and Microsystems (2020) 103413. doi: 10.1016/j.micpro.2020.103413.

[6] S. Luna, M. J. Pennock, Social media applications and emergency management: A literature review and research agenda. International Journal of Disaster Risk Reduction 28 (2018) 565–577. doi: 10.1016/j.ijdrr.2018.01.006.

[7] T. G. Penkova, A. V. Korobko, V. V. Nicheporchuk, L. F. Nozhenkova, On-line Control of the State of Technosphere and Environment Objects in Krasnoyarsk region, International Journal of Knowledge-based and Intelligent Engineering Systems 20(2) (2016) 65–74. doi: 10.3233/KES-160330.

[8] W. Inmon, Building the Data Warehouse, John Willey & Sons, NY, 1992.

[9] A. V. Chernov, V. A. Bogachev, E. V. Karpenko, M. A. Butakova, Y. V. Davidov, Rough and fuzzy sets approach for incident identification in railway infrastructure management system, in: XIX IEEE International Conference on Soft Computing and Measurements, 2016, pp. 228–230. doi: 10.1109/SCM.2016.7519736.

[10] C. Anderson, Creating a Data-Driven Organization, O'Reilly Media, (2015).

[11] V. V. Nicheporchuk, T. G. Penkova, Comprehensive analysis of territorial risk factors for estimating and managing the natural and technogenic safety, Issues of Risk Analysis 16 (4) (2019) 52–62. doi: 10.32686/1812-5220-2019-16-4-0-0.

[12] H. Wickham, G. Grolemund, R for Data Science: Import, Tidy, Transform, Visualize, and Model Data, O'Relly, 2016.