Available online at www.sciencedirect.com

ScienceDirect

Procedia Computer Science 112 (2017) 99–108

Procedia Computer Science

www.elsevier.com/locate/procedia

International Conference on Knowledge Based and Intelligent Information and Engineering Systems, KES2017, 6-8 September 2017, Marseille, France

Principal component analysis and cluster analysis for evaluating the natural and anthropogenic territory safety

T.G. Penkova[a,b,*]

[a]*Institute of Computational Modelling of the Siberian Branch of the Russian Academy of Sciences, 50/44 Akademgorodok, Krasnoyarsk, 660036, Russia*
[b]*Siberian Federal University, 79 Svobodny pr., Krasnoyarsk, 660041, Russia*

Abstract

This paper presents an approach to evaluating the natural and technogenic safety of the one of the largest regions in Siberia through the comprehensive analysis of territorial indicators in order to explore geographical variations and patterns in occurrence of emergencies by applying the data mining techniques – principal component analysis and cluster analysis – to data of the Territory Safety Passports. For data modeling, two principal components are selected and interpreted taking account of the contribution of the data attributes to the principal components. Data distribution on the principal components is analyzed at different levels of the territory detail: municipal areas and settlements. Two- and three- cluster structures are constructed in multidimensional data space; the main clusters features are investigated. The results of this analysis have allowed to identify the high-risk territories and rank them according to danger degree of occurrence of the natural and technogenic emergencies. This evaluation gives the basis for decision making and makes it possible for authorities to allocate the forces and means for territory protection more efficiently and develop a system of measures to prevent and mitigate the consequences of emergencies in the large region. The suggested in this work approach in terms of its stages, techniques and reasoning procedures can be considered as a model of comprehensive multidimensional analysis of the control objects in various areas.

Peer-review under responsibility of KES International

Keywords: comprehensive multidimensional analysis; principal component analysis; cluster analysis; data mining; terrritory safety evaluation; prevention of emergencies, territorial management.

* Corresponding author. Tel.: +7 913-516-9981; fax: +7-391-290-7476.
E-mail address: penkovatg@gmail.com

1877-0509
Peer-review under responsibility of KES International
10.1016/j.procs.2017.08.179

1. Introduction

Prevention of natural and technogenic emergencies is one of the major tasks of the territory management. Analytical support of decision-making processes based on modern technologies and efficient methods of data analysis is a necessary condition for improving the territorial safety system and management quality.

The Krasnoyarsk region is the second largest federal subject of Russia and the third largest subnational governing body by area in the world. The Krasnoyarsk region lies in the middle of Siberia and occupies an area of 2.4 million square kilometers, which is 13% of the country's total territory. This territory is characterised by heightened level of natural and technogenic emergencies which is determined by social-economic aspects, large resource potential, geographical location and climatic conditions[1]. In order to improve the population and territory safety, a lot of monitoring systems for on-line observation and for operational control of the state of technosphere and environment objects are being actively introduced within the region[2,3,4,5,6]. The Ministry of Emergency has enacted the structure and order of conducting the Territory Safety Passport, which defines a system of indicators to estimate the state of territory safety, the risk of emergencies and possible damages to create efficient prevention and mitigation actions[7]. At present, there are massive data collections about the state of controlled objects, occurred events and sources of emergencies[8,9]. However, we have to admit that the processing of stored data, aimed at obtaining the new and useful knowledge, is insufficient. The local databases remain unused, while the emergencies prediction, reasonable decisions and comprehensive analysis are sorely needed. Thus, identification of risk factors of emergencies based on monitoring data and investigation of their impact on key indicators of human safety are topical and important tasks in territorial management.

Data Mining, as the extraction of hidden predictive information from large databases, is a powerful modern technology of intelligent data processing. Data mining techniques provide the effective tool for discovering previously unknown, nontrivial, practically useful and interpreted knowledge needed to make decisions[10,11,12]. This paper presents an approach to evaluating the natural and technogenic safety of the one of the Krasnoyarsk region through the comprehensive analysis of territorial indicators in order to explore geographical variations and patterns in occurrence of emergencies by applying the data mining techniques – principal component analysis and cluster analysis – to data of the Territory Safety Passports.

The outline of this paper is as follows: Section 1 contains the introduction. Section 2 describes the initial data. Section 3 presents the principal component analysis: identification and interpretation of principal components; analysis of data distribution on the principal components at different levels of the territory detail. Section 4 presents the cluster analysis: construction of two- and three-cluster structures in multidimensional data space and analysis of their basic features. Section 5 draws the conclusion.

2. Data Description

Analysis of natural and technogenic safety indicators is based on data of the Territory Safety Passports of the Krasnoyarsk region for 2014 collected in Center of Emergency Monitoring and Prediction (CEMP). Original dataset contains 1,690 objects, essentially discrete settlements-level geographical entities of the Krasnoyarsk region, each with 12 measured attributes. Data attributes are listed in Table 1. One part of attributes characterizes the sensitivity of the territory to the risk factors effects (e.g. population density, the presence of industrial and engineering facilities) that is determined by the number of objects located on the territory (i.e. a number of potential sources of emergencies), it is so-called "object attributes". The other part of attributes characterizes the presence of potential factor that can damage the health of people, can cause irreversible damage to the environment that is determined by the statistic of events occurred in the territory (i.e. a number of emergencies), it is so-called "event attributes". In addition, some locational reference characteristics are used for data interpretation and map visualization.

The preliminary correlation analysis of original data has shown a fairly strong relationship between "object" and "event" attributes, therefore for further analysis we will consider the attributes that characterize population and events. The correlation coefficients are presented in Table 2.