

10. Коробко А.А., Ничепорчук В.В., Ноженков А.И. Динамическое формирование интерфейса ВЕБ-системы сбора данных мониторинга чрезвычайных ситуаций // Информатизация и связь. 2014. № 3. С. 59-64.

Creation of database of natural and technogenic safety indicators in the case of the Krasnoyarsk region

Valeriy Vasilyevich Nicheporchuk, PhD, Senior Researcher

This paper describes the major territorial management tasks and principles of forming database of natural and technogenic safety indicators. The process of data consolidation for comprehensive monitoring of settlements based on interdepartmental information exchange is demonstrated. Proposed structure of database and methods of data preparation and cartographic analysis are presented.

Keywords: danger event, database, spatial analysis, objects of protection and manage.

УДК 614.8:502/504:004.048

**АНАЛИЗ ПРИРОДНО-ТЕХНОГЕННОЙ БЕЗОПАСНОСТИ
НА ОСНОВЕ МЕТОДА ГЛАВНЫХ КОМПОНЕНТ
И КЛАСТЕРНОГО АНАЛИЗА
(НА ПРИМЕРЕ КРАСНОЯРСКОГО КРАЯ)**

Татьяна Геннадьевна Пенькова, к.т.н., старший научный сотрудник

Тел. 8 391 249 53 56, E-mail: penkova_t@icm.krasn.ru

Институт вычислительного моделирования СО РАН

http://icm.krasn.ru

Выполнен анализ характеристик природной и техногенной безопасности территорий Красноярского края, направленный на исследование географических особенностей и закономерностей возникновения ЧС на основе применения методов интеллектуальной обработки данных к данным паспортов безопасности населенных пунктов.

Ключевые слова: интеллектуальная обработка данных, анализ главных компонент, кластерный анализ, природно-техногенная безопасность.

Исследование выполнено при финансовой поддержке РФФИ в рамках научного проекта № 16-37-00014

Предупреждение чрезвычайных ситуаций (ЧС) природного и техногенного характера – одна из приоритетных задач органов территориального управления. Необходимым условием совершенствования системы территориальной безопасности является аналитическое обеспечение процессов принятия решений, основанное на применении современных технологий и эффективных методов анализа.

Территория Красноярского края характеризуется повышенным уровнем ЧС природного и техногенного происхождения [1]. С целью повышения безопасности населения и территорий края активно внедряются системы мониторинга за состоянием окружающей среды и объектов техносферы [2; 3; 4; 5; 6]. Приказом МЧС России утверждена структура и порядок ведения «Паспорта безопасности территории», который определяет систему показателей, позволяющих ор-



Т.Г. Пенькова

ганам власти оценивать состояние безопасности территорий и угрозу возникновения ЧС [7]. К настоящему времени в территориальных подразделениях МЧС России накоплены значительные объемы данных о состоянии объектов защиты, о событиях и источниках возникновения ЧС. Однако их обработка, направленная на получение новых и полезных знаний, недостаточная. Ведомственные базы данных остаются невостребованными, в то время как обоснованные управленческие решения, комплексный анализ и прогнозирование критических ситуаций становятся все более необходимыми. Выявление факторов риска возникновения ЧС по данным мониторинга и исследование их влияния на основные показатели безопасности жизнедеятельности населения и территории – актуальные и остро востребованные задачи территориального управления.

Методы интеллектуальной обработки предоставляют эффективный инструмент для обнаружения в данных ранее неизвестных, нетривиальных, практически полезных и доступных интерпретации знаний, необходимых для принятия решений [8]. В данной работе выполнен комплексный анализ характеристик природной и техногенной безопасности территорий Красноярского края, основанный на применении метода главных компонент и кластерного анализа к данным паспортов безопасности населенных пунктов. Результаты анализа позволяют определить зоны повышенного риска, ранжировать территории по степени опасности возникновения ЧС и более эффективно планировать и проводить мероприятия по предупреждению ЧС на территории Красноярского края.

Описание исходных данных

Анализ природно-техногенной безопасности территорий выполняется на основе данных паспортов безопасности населенных пунктов Красноярского края за 2014 г. Исходные данные представляют множество объектов и атрибутов. Объекты – населенные пункты Красноярского края; атрибуты – признаки, характеризующие состояние природно-техногенной безопасности и определяющие угрозу возникновения ЧС. Согласно данным паспортов безопасности [9], выделено 12 основных признаков. Часть из них характеризует количество расположенных на территории объектов (потенциальных источников возникновения ЧС) – «объектовые» признаки; другая часть признаков характеризует произошедшие на территории ЧС – «событийные» признаки. Описание признаков представлено в таблице 1.

Таблица 1

Описание признаков природно-техногенной безопасности территорий

№ п/п	Имя атрибута	Описание атрибута
1	Pop	Численность населения, чел.
2	Soc_object	Количество социально-значимых объектов (объекты образования, здравоохранения, соц. защиты, культуры, спорта)
3	Water_object	Количество опасных водных объектов
4	Indust_object	Количество потенциально-опасных промышленных объектов
5	Oil_line	Количество ниток нефтепровода (в радиусе 5 км от границ населенного пункта)
6	Munic_object	Количество объектов ЖКХ (объектов электроснабжения, объектов водоснабжения и теплоснабжения)
7	Food_event	Число подтоплений и затоплений (за последние 45 лет)
8	NFire_event	Число природных пожаров (в радиусе 5 км за последние 20 лет)
9	TFire_event	Число техногенных пожаров (за последние 20 лет)
10	Munic_event	Число аварий на объектах ЖКХ (за последние 15 лет)
11	Nat_event	Число природных событий за исключением природных пожаров, подтоплений и затоплений (за последние 15 лет)
12	Tech_event	Число техногенных аварий за исключением техногенных пожаров и аварий на ЖКХ (за последние 15 лет)

Предварительный корреляционный анализ исходных данных показал достаточно сильную зависимость между «объектовыми» и «событийными» признаками, поэтому в дальнейшем анализе рассматриваются только признаки, характеризующие численность

населения и события: Pop, Food_event, NFire_event, TFire_event, Munic_event, Nat_event, Tech_event. Помимо основных атрибутов, в обработке используются дополнительные характеристики – идентификаторы территорий, необходимые для визуализации объектов анализа. Итоговая таблица данных содержит 1690 объектов и 14 атрибутов (7 основных и 7 дополнительных). Анализ и визуализация многомерных данных выполняются с помощью средств системы «VidaExpert» [10]. Визуализация данных на географических картах выполняется с помощью средств картографирования «ArcGIS» [11].

Анализ главных компонент

Метод главных компонент (Principal Component Analysis) – один из широко используемых методов Data Mining, направленный на изучение закономерностей в структуре многомерных данных – один из эффективных способов понижения размерности, состоящий в переходе к новому ортогональному базису, оси которого ориентированы по направлениям максимальной дисперсии набора входных данных [12].

Одной из первоочередных задач, возникающих при использовании метода главных компонент, является оценка числа главных компонент. В общем случае, метод дает возможность по k исходным признакам выделить k компонент. В таблице 2 представлены результаты расчета собственных векторов ковариационной матрицы, расположенные в порядке убывания собственных значений.

Таблица 2

Результаты расчета главных компонент

Компоненты	1	2	3	4	5	6	7
Собственные значения	0,40357	0,24861	0,14113	0,11624	0,075068	0,010358	0,00502
Накопленная дисперсия	0,50357	0,65218	0,79331	0,90955	0,98462	0,99498	1
Pop	0,50913	0,10888	0,11149	0,11281	0,22748	0,18226	0,78665
TFire_event	0,51347	0,08304	0,06051	0,08742	0,17071	0,61552	-0,55691
NFire_event	0,06017	0,43939	-0,87581	0,18588	-0,02195	-0,03308	0,01172
Munic_event	0,50327	0,09611	0,11964	0,08378	0,25051	-0,76422	-0,26337
Food_event	0,23478	-0,31425	-0,3248	-0,85321	0,10858	-0,00384	0,02943
Nat_event	0,08601	-0,82162	-0,31094	0,45818	0,10299	-0,01485	0,01008
Tech_event	0,39714	-0,0716	0,01927	0,01319	-0,91295	-0,05053	0,02396

На основе правила Кайзера (Kaiser's rule) и правила сломанной трости (Broken-stick model) [13], для исходного набора данных определены две главные компоненты (PC1 и PC2), описывающие 65% дисперсии. Характер главных компонент определяется весами признаков (собственными значениями). Вклад признаков в первую и вторую главные компоненты показан в виде графиков на рис. 1.

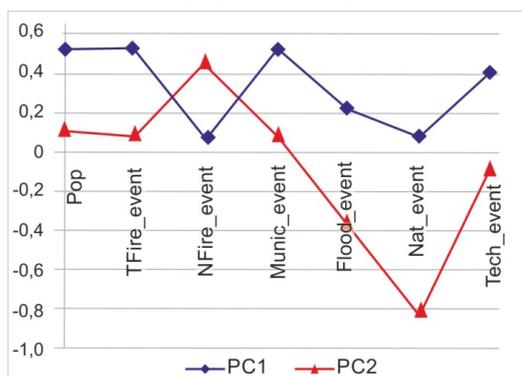


Рис. 1. Вклад атрибутов в первую (PC1) и вторую (PC2) главную компоненту

Первая компонента характеризуется высокой численностью населения, высоким уровнем техногенных пожаров, аварий на ЖКХ и прочими техногенными ЧС, а также незначительным уровнем событий природного характера. В совокупности, первая компонента определяет крупные населенные пункты с высоким уровнем техногенной опасности. Вторая компонента характеризуется низкой численностью населения, высоким уровнем природных пожаров и отсутствием других природных ЧС, а также невысоким уровнем событий техногенного характера. В совокупности вторая компонента определяет небольшие населенные пункты с высоким уровнем природных пожаров. Характер первой и второй главной компоненты позволяет сделать вывод, что для Красноярского края наибольшую угрозу представляют техногенные и природные пожары.

Распределение проекций на первую и вторую главные компоненты на карте представлено на рис. 2 и 3 соответственно. Отрицательные значения проекций (в диапазоне [-1; 0]) соответствуют группе 1 – синий цвет, наиболее высокие положительные значения проекций (в диапазоне (0,5;1]) соответствуют группе 2 – красный цвет.

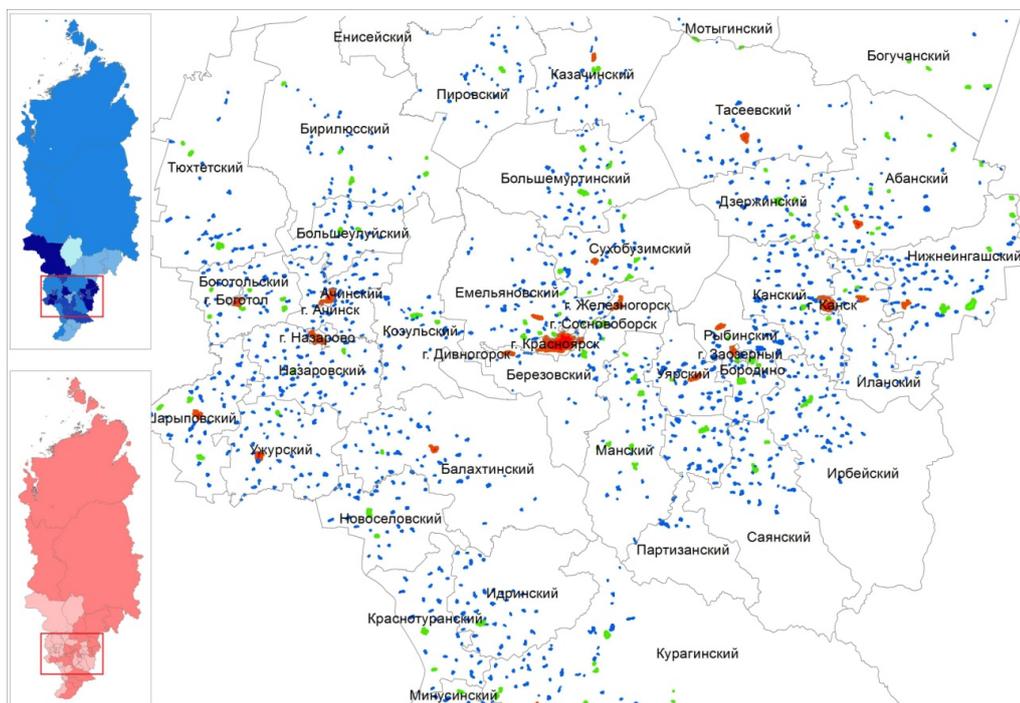


Рис. 2. Распределение проекций на первую главную компоненту в разрезе муниципальных образований и населенных пунктов края

Наиболее низкие значения проекций на первую главную компоненту (рисунок 2, синий цвет) наблюдаются для населённых пунктов: Усть-Камо, Шигашет, Касово, Верхнекемское, Средняя Шилка, Комаровский, Новый Сатыш, Ангутиха, Лебедь. На территории данных населённых пунктов нет ни одного социально-значимого объекта и не зарегистрировано ни одного жителя. Отсутствие хозяйственной деятельности объясняет низкий уровень (отсутствие) техногенных пожаров. Наиболее высокие значения проекций на первую главную компоненту (рисунок 2, красный цвет) наблюдаются для населённых пунктов: Красноярск, Норильск, Ачинск, Канск, Минусинск, Лесосибирск, Назарово, Емельяново, Абан, Енисейск, где численность населения и количество социальных и промышленных объектов выше средних значений по краю.

Наиболее низкие значения проекций на вторую главную компоненту (рисунок 3, синий цвет) наблюдаются для населённых пунктов: Туруханск, Черемшанка, Танзыбей, Емельяново, Ермаковское, Нижний Ингаш, Вельмо, Курагино, Сухобузимское, Ужур. Низкий уровень природных пожаров объясняется двумя моментами: отсутствием растительности (степные районы Западной и Южной группы) и отсутствием поселений в лесной зоне (Эвенкийский автономный округ, Енисейский и Туруханский районы). Наиболее высокие значения проекций на вторую главную компоненту (рисунок 3,

красный цвет) наблюдаются для населённых пунктов: Старцево, Тиличеть, Курай, Крутояр, Байкал, Уджей, Глинный, Абалаково, Проточный. Высокий риск природных пожаров наблюдается в крупных поселениях, прилегающих к лесным массивам. Кроме того, вероятность возникновения природных пожаров сохраняется и в крупных городах из-за наличия лесных массивов в их границах, однако переход очага в крупный пожар маловероятен за счёт проведения профилактических мероприятий.

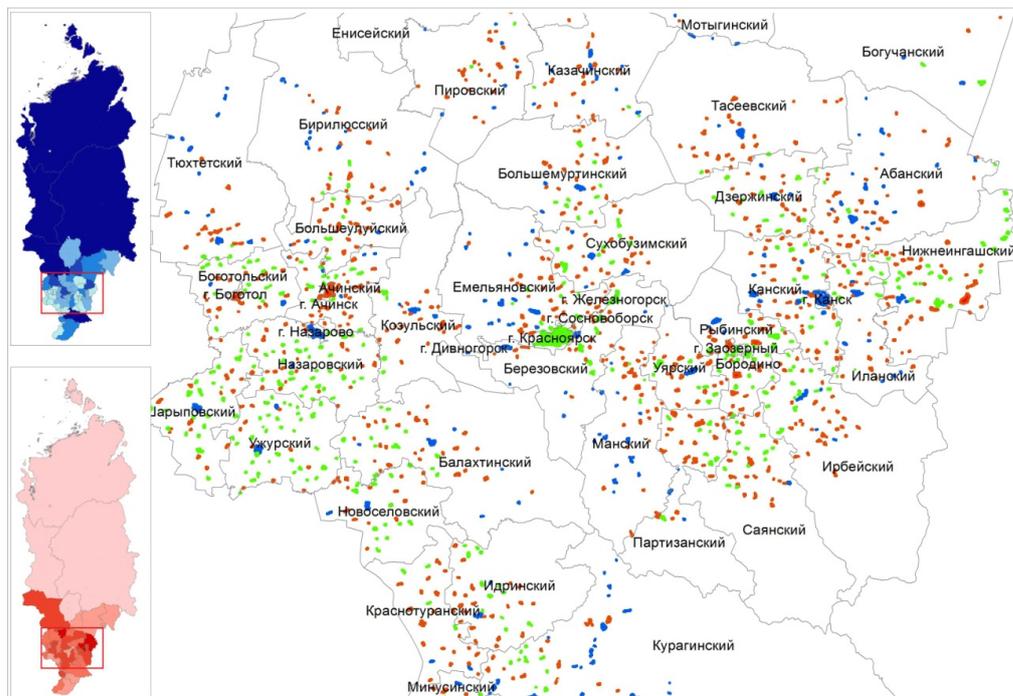


Рис. 3. Распределение проекций на вторую главную компоненту в разрезе муниципальных образований и населенных пунктов края

Кластерный анализ

Кластерный анализ (Cluster Analysis) – метод классификации, основанный на обнаружении скрытых зависимостей и выявлении структуры в данных. Метод заключается в объединении объектов в непересекающиеся группы – кластеры на основе близости значений их атрибутов [14]. Наиболее распространенным алгоритмом кластеризации является алгоритм k -средних (k -means). Алгоритм разбивает множество элементов векторного пространства на заранее известное число кластеров k , путем минимизации суммарного квадратичного отклонения точек кластеров от их центров. Особенностью реализации алгоритма k -средних является наличие гипотезы о количестве кластеров. Анализируемые данные характеризуются трёхкластерной структурой, что определяется наличием двух главных компонент.

Двухкластерная структура ($k = 2$) формируется первой главной компонентой: кластер 1 (синий цвет) содержит 352 объекта, кластер 2 (красный цвет) содержит 1338 объектов. Характеристика кластеров определяется по вкладу атрибутов. На рис. 4 представлено распределение данных по атрибутам в двухкластерной структуре.

Из рис. 4 видно, что два кластера существенно отличаются по численности населения и числу техногенных пожаров. Кроме того, для кластера 1, в отличие от кластера 2, более характерны аварии на объектах ЖКХ и в меньшей степени характерны природные пожары и наводнения. В совокупности кластер 1 объединяет крупные населённые пункты с развитой инфраструктурой, где присутствует угроза ЧС техногенного характера, и с богатой природной средой (лесные массивы, водные объекты), которая определяет возможность возникновения ЧС природного характера. В свою очередь, кластер 2 объединяет мелкие населённые пункты с минимальным риском возникнове-

ния ЧС как природного, так и техногенного характера. Результаты кластеризации на карте представлены на рис. 5.

Наибольшее количество объектов кластера 1 наблюдается в Курагинском, Ужурском, Боготольском, Назаровском районах, наибольшее количество объектов кластера 2 наблюдается в Енисейском, Абанском, Нижнеингашском, Канском и Емельяновском районах. Среди крупных населённых пунктов к представителям кластера 1 относятся: Красноярск, Норильск, Ачинск, Канск, Железногорск, Зеленогорск, Минусинск, Лесосибирск, Сосновоборск, Назарово, Шарыпово; к представителям кластера 2 относятся: Новохайский, Солнечный, Козулька, Северо-Енисейский, Подгорный, Краснотуранск, Зыково, Подтесово, Кедрово, Краснокаменск. Угроза возникновения ЧС техногенного характера в крупных населённых пунктах связана с наличием производственных объектов, объектов ЖКХ с высокой степенью износа оборудования. Высокий риск затопления населённых пунктов и инфраструктуры, при обеспеченности паводков 10% и менее, связан с интенсивной застройкой в поймах рек и местах с высоким уровнем грунтовых вод.

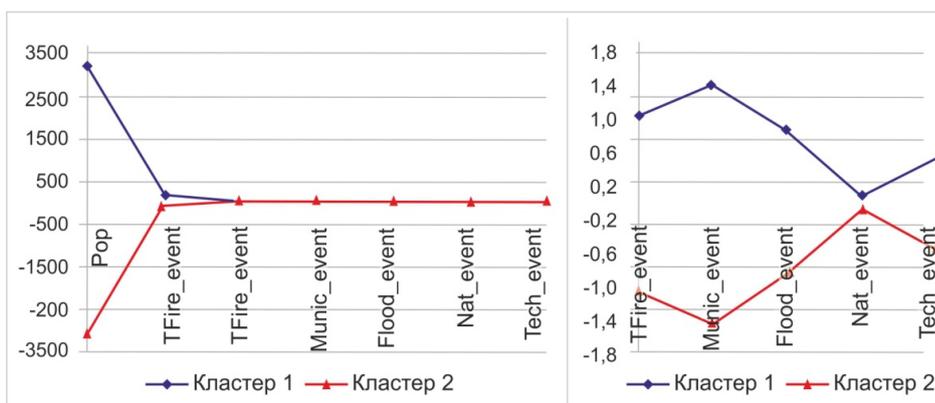


Рис. 4. Распределение данных по атрибутам в двухкластерной структуре

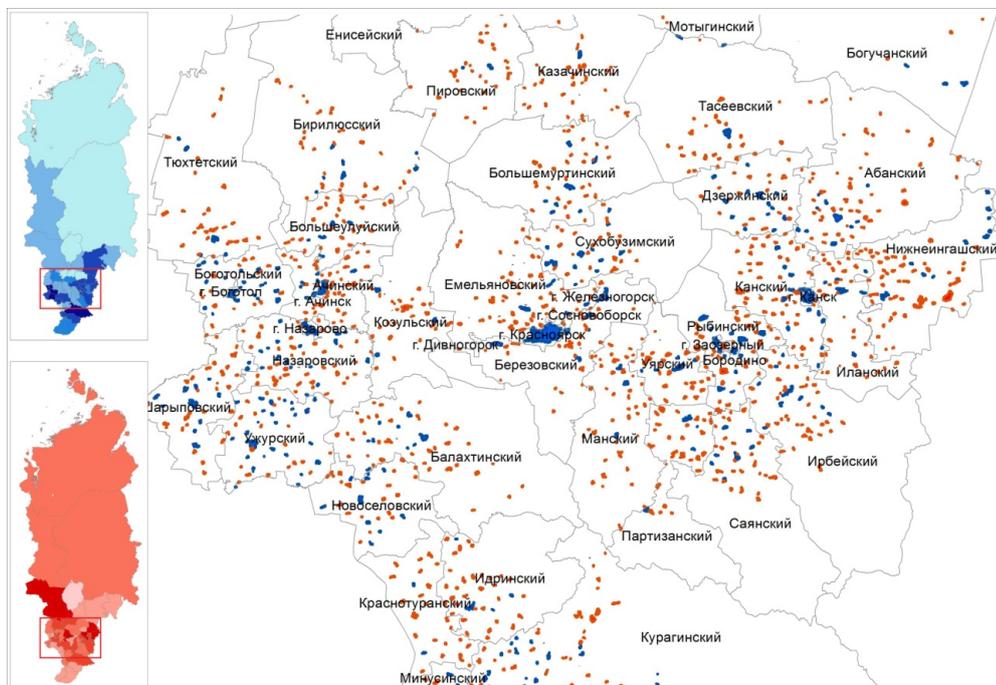


Рис. 5. Двухкластерная структура данных в разрезе муниципальных образований и населенных пунктов края

Трехкластерная структура ($k = 3$) формируется второй главной компонентой: кластер 1 (синий цвет) содержит 80 объектов, кластер 2 (красный цвет) содержит 719 объектов, кластер 3 (зеленый цвет) содержит 890 объектов. На рис. 6 представлено распределение данных по атрибутам в трехкластерной структуре.

Из рис. 6 видно, что кластер 1 существенно отличается от кластера 2 и кластера 3 и характеризуется большой численностью населения, большим числом техногенных пожаров, аварий на ЖКХ и других событий техногенного характера. Кластер 2 и кластер 3 характеризуются небольшой численностью населения и низким уровнем событий природного и техногенного характера. При этом кластер 3 отличается более высоким уровнем природных пожаров. Следовательно, кластер 1 объединяет более крупные населённые пункты с развитой инфраструктурой, где присутствует угроза техногенного характера. Кластер 2 объединяет населённые пункты с минимальным риском возникновения ЧС природного и техногенного характера. Кластер 3 объединяет населённые пункты, где основная угроза – природные пожары. Результаты кластеризации на карте представлены на рис. 7.

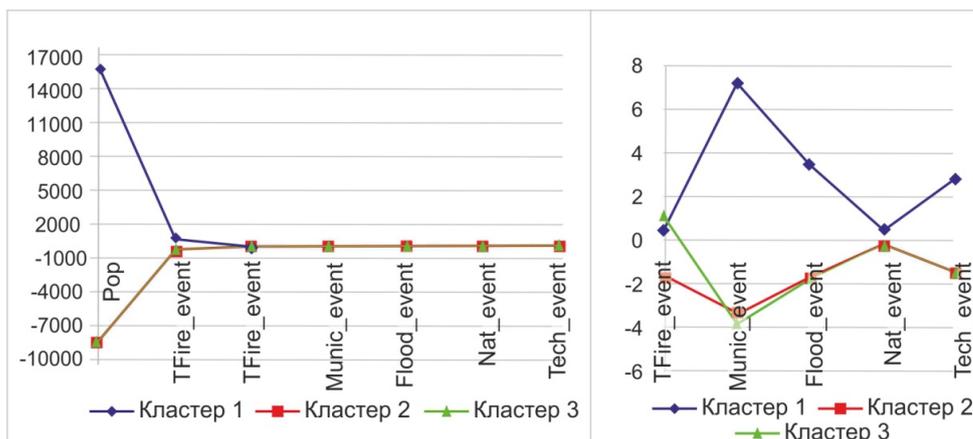


Рис. 6. Распределение данных по атрибутам в трёхкластерной структуре

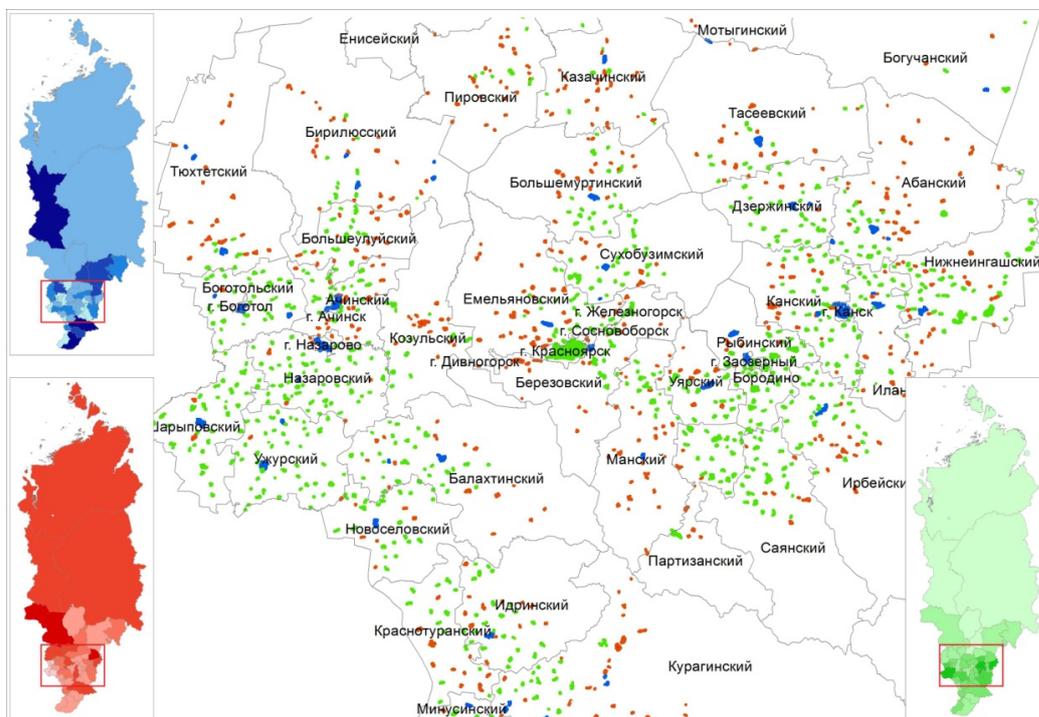


Рис. 7. Трёхкластерная структура данных в разрезе муниципальных образований и населённых пунктов края

Наибольшее количество объектов кластера 1 наблюдается в Туруханском, Ермаковском, Каратузском, Курагинском районах, наибольшее количество объектов кластера 2 наблюдается в Абанском, Енисейском, Емельяновском, Пировском, Курагинском районах, наибольшее количество объектов кластера 3 наблюдается в Назаровском, Ужурском, Канском, Нижнеингашском, Рыбинском районах. Среди крупных населённых

ных пунктов к представителям кластера 1 относятся: Ачинск, Канск, Зеленогорск, Лесосибирск, Минусинск, Шарыпово, Назарово, Норильск; к представителям кластера 2 относятся населённые пункты: Дивногорск, Козулька, Северо-Енисейский, Подгорный, Краснотуранск, Кедровый, Кошурниково, Верхнепашино, Байкит; к представителям кластера 3 относятся населённые пункты: Красноярск, Железногорск, Сосновоборск, Бородино, Шушенское, Кодинск, Агинское. Высокий риск природных пожаров наблюдается в небольших населённых пунктах, непосредственно прилегающих к лесным массивам, а также в крупных населённых пунктах, в границах которых расположены лесные массивы.

Заключение

Автор считает, что в данной работе новыми являются следующие положения и результаты. На основе методов интеллектуальной обработки данных по данным паспортов безопасности впервые выполнен комплексный анализ характеристик природной и техногенной безопасности Красноярского края с уровнем детализации территорий до населённых пунктов. Выполнена оценка главных компонент и их интерпретация с учётом вклада анализируемых характеристик. Проанализировано распределение данных по главным компонентам на разных уровнях детализации территории: группы районов, муниципальные образования и населённые пункты. В многомерном пространстве данных построены двух- и трёхкластерные структуры.

Результаты анализа показали, что наибольшую угрозу территориям Красноярского края представляют техногенные и лесные пожары. Высокий риск техногенных пожаров наблюдается в крупных населённых пунктах, в которых численность населения, количество социально-значимых и промышленных объектов выше средних значений. Высокий риск природных пожаров наблюдается в крупных поселениях, непосредственно прилегающих к лесным массивам, а также в городах, в границах которых расположены лесные массивы. Выявленные географические особенности и закономерности позволили определить зоны повышенного риска, ранжировать территории и населённые пункты по степени опасности возникновения ЧС природного и техногенного характера. Результаты исследования дают возможность более эффективно планировать и проводить стратегические мероприятия по предупреждению и смягчению последствий ЧС на территории Красноярского края.

Литература

1. Государственный доклад «О состоянии защиты населения и территорий Красноярского края от чрезвычайных ситуаций природного и техногенного характера в 2014 году» – Красноярск: Главное управление МЧС РФ по Красноярскому краю, 2015. 260 с.
2. Об утверждении положения о краевой подсистеме мониторинга, лабораторного контроля и прогнозирования чрезвычайных ситуаций природного и техногенного характера на территории Красноярского края: Постановление Правительства Красноярского края от 9.02.2011 г. № 80-п.
3. *Penkova T.G., Korobko A.V., Nicheporchuk V.V., Nozhenkova L.F.* On-line modelling and assessment of the state of technosphere and environment objects based on monitoring data // *Procedia Computer Science: Elsevier.* 2014. Vol. 35. PP. 156-165.
4. *Yronen Y.P., Yronen E.A., Ivanov V.V., Kovalev I.V., Zelenkov P.V.* The concept of creation of information system for environmental monitoring based on modern gis-technologies and earth remote sensing data // *IOP Conf. Series: Materials Science and Engineering.* 2015. V. 94. 012023.
5. *Шапарев Н.Я.* Мониторинг окружающей среды Красноярского края в показателях устойчивого природопользования // *Горный информационно-аналитический бюллетень (научно-технический журнал).* 2009. Т. 18. № 12. С. 110-113.
6. *Брюханова Е.А., Кобалинский М.В., Сибгатулин В.Г., Шишацкий Н.Г.* Совершенствование информационного обеспечения экологического мониторинга как инструмента устойчивого социально-экономического развития региона (на примере Красноярского края) // *Информатизация и связь.* 2014. № 1. С. 43-47.

7. Об утверждении типового паспорта безопасности территорий субъектов Российской Федерации и муниципальных образований: Приказ МЧС России от 25.10.2004 г. №484 (ред. от 11.09.2013).
8. Williams G.J., Simoff S.J. Data Mining: Theory, Methodology, Techniques, and Applications: Springer. 2006. 329 p.
9. Ничепорчук В.В., Пенькова Т.Г. Паспорт территорий – динамический инструмент анализа опасностей // Научные и образовательные проблемы гражданской защиты. 2014. № 1. С. 3-8.
10. Gorban A., Pitenko A., Zinovyev A. ViDaExpert: user-friendly tool for nonlinear visualization and analysis of multidimensional vectorial data: Cornell University Library. 2014 URL: <http://arxiv.org/abs/1406.5550>.
11. Using ArcViewGIS: The Geographic Information System of Everyone: ESRI Press. – 1999. 350 p.
12. Abdi H., Williams L. Principal Components Analysis. Wiley Interdisciplinary Reviews: Computational Statistics. 2010. N. 2(4). PP. 439-459.
13. Peres-Neto P., Jackson D., Somers K. How many principal components? Stopping rules for determining the number of non-trivial axes revisited // Computational Statistics & Data Analysis. – 2005. N. 49(4). PP. 974-997.
14. Jain A., Dubes R. Algorithms for Clustering Data. Michigan State University: Prentice Hall, 1988. 320 p.

Analysis of natural and technogenic safety based on principal component analysis and cluster analysis techniques (in case of Krasnoyarsk region)

Penkova Tatiana Gennadyevna, PhD, Senior Research Associate

This paper presents a comprehensive analysis of natural and technogenic safety indicators of the Krasnoyarsk region in order to explore geographical variations and patterns in occurrence of emergencies by applying the multidimensional analysis techniques – principal component analysis and cluster analysis – to data of the Territory Safety Passports.

Keywords: data mining, principal component analysis, cluster analysis, natural and technogenic safety

УДК 004.42

**ОБРАБОТКА МНОГОМЕРНЫХ ДАННЫХ
СРЕДСТВАМИ ТЕХНОЛОГИИ ТОФИ**

*Сауле Кайроллиевна Сагнаева, к.ф.-м.н., доцент,
Евразийский национальный университет им.Л.Н.Гумилева
Тел.: +7 705 709 9703, e-mail: sagnaeva_tar@mail.ru,*

*Шынар Ерланкызы, магистрант
Евразийский национальный университет им.Л.Н.Гумилева
Тел.: +77476917300, e-mail: shinara.92@mail.ru
fit.enu.kz*

В статье рассматриваются вопросы анализа и визуализации многомерных данных средствами технологии ТОФИ. Описаны технологические особенности построения хранилища данных ТОФИ, многомерных кубов и срезов кубов. Основное назначение технологии ТОФИ – поддержка аналитической деятельности, произвольных запросов пользователей-аналитиков.