

УДК: 004.896

DOI: 10.25559/SITITO.14.201803.679-685

ИНТЕЛЛЕКТУАЛЬНЫЙ АНАЛИЗ БОЛЬШИХ ПРОСТРАНСТВЕННО-ВРЕМЕННЫХ ДАННЫХ ДЛЯ СЛУЖБ ЭКСТРЕННОГО РЕАГИРОВАНИЯ

А.Р. Гараева¹, Р.Н. Минниханов², М.В. Дагаева², С.С. Кильдеева¹, И.В. Аникин¹

¹ Казанский национальный исследовательский технический университет им. А. Н. Туполева – КАИ, г. Казань, Россия

² ГБУ «Безопасность дорожного движения», г. Казань, Россия

INTELLIGENT ANALYSIS OF LARGE SPATIAL-TEMPORAL DATA FOR EMERGENCY SERVICES

Alina R. Garaeva¹, Rifkat N. Minnikhanov², Maria V. Dagaeva², Sophya S. Kildeeva¹, Igor V. Anikin¹

¹ Kazan National Research Technical University named after A. N. Tupolev – KAI, Kazan, Russia

² State Budget Institution «Road Traffic Safety», Kazan, Russia

© Гараева А.Р., Минниханов Р.Н., Дагаева М.В., Кильдеева С.С., Аникин И.В., 2018

Ключевые слова

Системы управления чрезвычайными ситуациями; пространственно-временные данные; интеллектуальный анализ пространственных данных; интеллектуальный анализ расположения пространственных шаблонов; выявление пространственных аномалий; анализ пространственной автокорреляции.

Аннотация

В последнее время аварийно-спасательные службы проявляют большой интерес к использованию геоинформационных систем (ГИС) для повышения эффективности мониторинга чрезвычайных событий и быстрого реагирования. Эффективность использования таких систем можно повысить путем применения в них методов интеллектуального анализа пространственно-временных данных с целью обеспечения поддержки принятия решений и извлечения потенциально полезных знаний, которые могли бы помочь эффективно обнаруживать чрезвычайные ситуации в режиме реального времени и предотвращать некоторые происшествия.

Пространственно-временные данные играют большую роль в различных областях науки, таких как география, экология, здравоохранение, безопасность. На сегодняшний день рост населения и повышение уровня социальной активности имеет значительное воздействие на городскую среду и ведет к возникновению чрезвычайных ситуаций, таких как пожары, преступления, угрозы террористических актов, дорожно-транспортные происшествия. В связи с этим государственные органы нуждаются в новых решениях для эффективного реагирования и мониторинга чрезвычайных происшествий. Однако хранение и обработка больших объемов пространственно-временных данных на текущий момент все еще является серьезной проблемой, с которой сталкиваются службы, связанные с реагированием на чрезвычайные ситуации. В рамках данной работы был рассмотрен ряд подходов к анализу пространственных данных: анализ пространственных шаблонов, обнаружение пространственно-временных аномалий и анализ пространственной автокорреляции.

Об авторах:

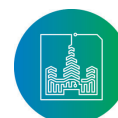
Гараева Алина Раисовна, аспирант, Казанский национальный исследовательский технический университет им. А. Н. Туполева – КАИ (420111, Россия, г. Казань, ул. К. Маркса, д. 10), ORCID: <http://orcid.org/0000-0002-3516-1424>, crazyalin91@mail.ru

Минниханов Рифкат Нурғалиевич, доктор технических наук, член-корреспондент АН Республики Татарстан, директор, ГБУ «Безопасность дорожного движения» (420000, Россия, г. Казань, ул. Оренбургский тракт, д. 5), ORCID: <http://orcid.org/0000-0001-9166-2955>, Priemnaa.Minnihanov@tatar.ru

Дагаева Мария Витальевна, главный специалист Центра интеллектуальных транспортных систем, ГБУ «Безопасность дорожного движения» (420000, Россия, г. Казань, ул. Оренбургский тракт, д. 5), ORCID: <http://orcid.org/0000-0002-5444-9669>

Кильдеева Софья Сергеевна, магистр, Казанский национальный исследовательский технический университет им. А. Н. Туполева – КАИ (420111, Россия, г. Казань, ул. К. Маркса, д. 10), ORCID: <http://orcid.org/0000-0001-8871-2632>

Аникин Игорь Вячеславович, доктор технических наук, доцент, заведующий кафедрой систем информационной безопасности, Казанский национальный исследовательский технический университет им. А. Н. Туполева – КАИ (420111, Россия, г. Казань, ул. К. Маркса, д. 10), ORCID: <http://orcid.org/0000-0001-9478-4894>



Keywords

Emergency management systems; space-time data; intellectual analysis of spatial data; intellectual analysis of the location of spatial patterns; identification of spatial anomalies; spatial autocorrelation analysis.

Abstract

Recently, rescue services have shown great interest in the use of geographic information systems (GIS) to improve the efficiency of monitoring emergency events and rapid response. The efficiency of using such systems can be enhanced by applying spatial-temporal data mining techniques to provide decision support and extracting potentially useful knowledge that could help effectively detect emergencies in real time and prevent some incidents. Spatio-temporal data play a large role in various fields of science, such as geography, ecology, health, safety. Today, population growth and increased social activity has a significant impact on the urban environment and leads to emergencies such as fires, crimes, threats of terrorist acts, and road accidents. In this regard, government agencies need new solutions to effectively respond and monitor emergencies. However, the storage and processing of large amounts of spatio-temporal data is currently still a serious problem faced by services related to emergency response. As part of this work, a number of approaches to the analysis of spatial data were considered: the analysis of spatial patterns, the detection of space-time anomalies and the analysis of spatial autocorrelation.

Введение

На сегодняшний день информационные системы управления чрезвычайными ситуациями могут быть усовершенствованы включением в них инструментов обработки больших данных для обеспечения эффективного хранения, а также для внедрения функционала поддержки принятия решений. Кроме того, анализ исторических данных о чрезвычайных ситуациях методами интеллектуального анализа данных позволит извлекать новые потенциально полезные знания, которые могут помочь эффективно обнаруживать и предотвращать чрезвычайные ситуации в режиме реального времени.

Пространственный анализ данных [1] относится к категории задач Data Mining, и представляет собой процесс получения скрытых знаний, пространственных отношений и интересных пространственных шаблонов. Для решения задач управления чрезвычайными ситуациями можно выделить следующие подходы к интеллектуальному анализу пространственно-временных данных: 1) пространственная кластеризация; 2) анализ пространственных шаблонов; 3) обнаружение пространственно-временных аномалий; 4) анализ пространственной автокорреляции. Кроме того, такие данные принадлежат к категории больших данных, что влечет необходимость их распределенного хранения и обработки. Однако, стек современных технологий, разработанный для обработки больших данных, таких как Apache Hadoop и Apache Spark, не поддерживают встроенных функций для анализа пространственно-временных данных, что затрудняет их практическое применение. В данной статье предлагаются методы, позволяющие решить ряд задач интеллектуального анализа пространственно-временных данных применительно к мониторингу и управлению чрезвычайными происшествиями. Дано описание разработанного программного комплекса, реализующего рассмотренные методы, и представлены архитектурные и эксплуатационные характеристики разработанной системы.

1. Интеллектуальный анализ пространственно-временных данных для управления чрезвычайными ситуациями**1.1. Интеллектуальный анализ пространственных шаблонов**

Анализ пространственных шаблонов (англ. spatial collocation pattern) в основном направлен на поиск интересных свя-

зей между пространственными объектами и позволяет ответить на следующий вопрос: имеют ли некоторые пространственные объекты тенденцию случаться и находиться в непосредственной географической близости [2, 15]. Термин «пространственный шаблон» можно определить, как набор пространственных объектов, часто наблюдаемых в пространственной близости. В разрезе вопроса управления чрезвычайными ситуациями пространственные шаблоны могут быть определены как подмножество чрезвычайных событий, с высокой вероятностью происходящих в одно время и близко располагающихся друг к другу в пространстве. Такие шаблоны формально определяются в виде ассоциативных правил и могут быть интерпретированы как сценарии развития чрезвычайных ситуаций:

[пожар] => [пострадавшие], вероятность 0.80%;

[автомобильная авария] => [пострадавшие], вероятность 0.77%.

Для анализа пространственных шаблонов в основном используются следующие подходы: применение пространственной статистики и применение подходов интеллектуального анализа данных. Методы пространственной статистики основаны на вычислении пространственной корреляции, которая показывает зависимость между различными атрибутами пространственных объектов [3]. Методы интеллектуального анализа данных основаны на двух основных подходах: кластерный анализ [4] и анализ ассоциативных правил [5]. Формально процесс обнаружения пространственных шаблонов состоит из двух шагов: определения набора транзакций и анализа ассоциативных правил. В рамках данной работы для определения набора транзакций был использован алгоритм кластеризации DBSCAN (Density Based Spatial Clustering of Applications with Noise) [6]. Основная идея DBSCAN заключается в кластеризации пространственных объектов на основе плотности распределения в пространстве. Одним из основных его преимуществ является возможность эффективной обработки больших пространственных данных [6].

После определения набора транзакций может быть применен алгоритм интеллектуального анализа ассоциативных правил для определения пространственных шаблонов. Одним из первых в качестве алгоритма анализа ассоциативных правил был предложен алгоритм Apriori. Позднее был предложен алгоритм FP-growth, результативность которого намного выше, чем у классического алго-



ритма Apriori. Более того, библиотека машинного обучения Spark MLlib реализует алгоритм анализа ассоциативных правил- FP-growth [7], который принимает на вход набор транзакций и два параметра: minSupport (порог для определения частоты появления элемента в выборке) и numPartitions (уровень деления набора дан-

ных для распределенной работы). Дополнительно устанавливается порог доверительного значения для идентификации набора ассоциативных правил, чтобы устранить бессмысленные правила с низкой достоверностью. На рисунке 1 схематично изображен процесс интеллектуального анализа пространственных шаблонов.

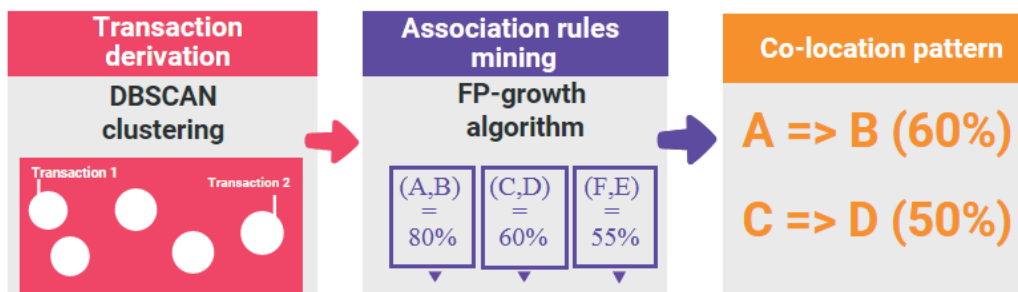


Рис. 1. Процесс интеллектуального анализа пространственных шаблонов
Fig. 1. Spatial Pattern Mining Process

1.2. Обнаружение пространственно-временных аномалий

Обнаружение аномалий широко используется для выявления отклоняющегося от нормы поведения во многих областях, таких как мошенничество с кредитными картами, мониторинг окружающей среды, контроль событий криминального характера, медицинская помощь и так далее. Согласно классическому определению, аномалия - это выявление объекта из набора данных, который существенно отличается от остальных объектов [1]. Пространственно-временные аномалии имеют свои особенности, которые необходимо учитывать: географическое положение и время. Пространственно-временная аномалия [11] - это объект наблюдаемой выборки, значения атрибутов которого слишком отличаются от значений атрибутов других пространственно-временных объектов в пространственно-временной близости. Большинство существующих алгоритмов используют только временные или пространственные аспекты для обнаружения аномалий. В рамках данной работы для выявления временных аномалий использован метод прогнозирования временных рядов с помощью статистической модели ARIMA

ARIMA (autoregressive integrated moving average model) - обобщение модели ARMA, используемая для дифференцирования частных значений нестационарного временного ряда. ARIMA (p, q, d), где p - количество наблюдений, d - степень дифференцирования, q - размер скользящего окна. В части AR (autoregression) используется зависимое отношение между наблюдением и количеством запаздывающих наблюдений. I (integrated) - используется для формирования стационарных временных рядов. MA (moving average) - зависимость между наблюдением и остаточной ошибкой от модели скользящего среднего, применяемой к запаздывающим наблюдениям [8].

В качестве альтернативного метода для обнаружения пространственно-временных аномалий используется рекуррентная нейронная сеть (RNN). Рекуррентная нейронная сеть эффективна для задач прогнозирования благодаря сбору предыдущих выходных данных, так как зависит от предыдущих вычислений. Кроме того, RNN [18] имеет "память", которая хранит информацию о том, что было вычислено ранее. «Длиительные зависимости» - это отношения между событиями, разделенными большим количеством моментов времени [9]. Обратное

распространение во времени, исчезающие градиенты, единицы долговременной и кратковременной памяти являются примерами RNN. LSTM это тип RNN и основная его особенность заключается в том, что он более достоверен, чем RNN, в отношении больших временных последовательностей, и исключает долгосрочной проблемы зависимости. LSTM содержит блоки памяти в скрытом слое. В свою очередь блоки памяти также содержат ячейки памяти и слои. Ячейки памяти отвечают за хранение временного состояния сети, а слои управляют хранением или удалением информации. Существует 3 типа слоев [10]:

1. Входной слой (управляет новыми входными данными, позволяет добавлять новый входные данные или нет);
2. "Забывающий" слой (удаляет информацию);
3. Выходной слой (отвечает за значение, которое используется для расчета выхода LSTM. LSTM успешно используется для прогнозирования временных рядов) [10].

1.3. Анализ пространственной автокорреляции

Пространственная автокорреляция является одним из основных задач пространственной статистики. Пространственная автокорреляция позволяет обнаружить определенные отношения между пространственными объектами и определить, зависит ли распределение пространственных объектов друг от друга или распределение имеет случайный характер [11]. Выделяют два вида пространственной автокорреляции: позитивную и негативную.

Для вычисления пространственной автокорреляции в первую очередь необходимо определить, что подразумевается под двумя наблюдениями, находящимися близко друг к другу, то есть должна быть произведена оценка расстояния между ними. Эти расстояния представляются в пространственной матрице весов, которая определяет отношения между местоположениями, где были сделаны измерения. Если данные собираются в виде географических координат, то матрица весов будет иметь нули по диагонали. После того, как пространственная матрица весов определена, пространственная автокорреляция может быть измерена индексом Морана I, статистикой Getis-Ords G, статистикой Geary's C.



2. Программный комплекс интеллектуального анализа пространственно-временных данных

Выше описанные методы были реализованы в программном комплексе анализа больших пространственно-временных данных для анализа сведений о чрезвычайных ситуациях. В качестве основных платформ для разработки с целью обеспечения высокой вычислительной производительности были выбраны Apache Hadoop и Apache Spark.

2.1. Процесс интеллектуального анализа пространственных шаблонов

Основной процесс интеллектуального анализа пространственных шаблонов изображен на рисунке 2. Он начинается с загрузки данных из распределенной файловой системы Hadoop HDFS. Информация о чрезвычайных событиях (id, местоположение, время) передается алгоритму кластеризации DBSCAN. Следующим шагом в результате этого шага получается множество транзакций-кластеров (событий, находящихся в непосредственной близости во времени и пространстве). После того, как набор транзакций определен, полученный набор может использоваться в качестве входных данных для алгоритма FP-Growth и поиска ассоциативных правил. Пример полученных пространственных шаблонов в наборе данных чрезвычайных ситуаций представлен на рисунке 3.

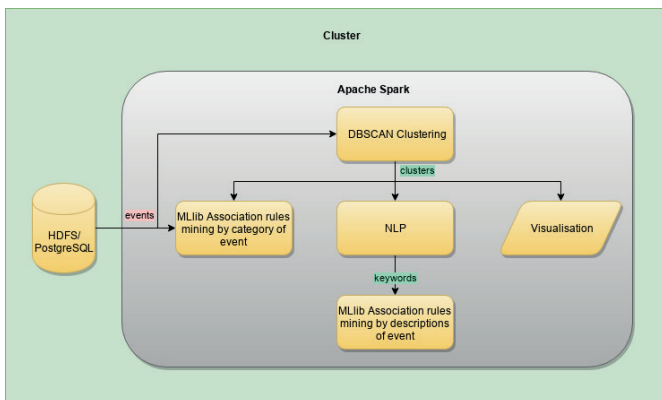


Рис. 2. Процесс интеллектуального анализа пространственных шаблонов
Fig. 2. Spatial Pattern Mining Process

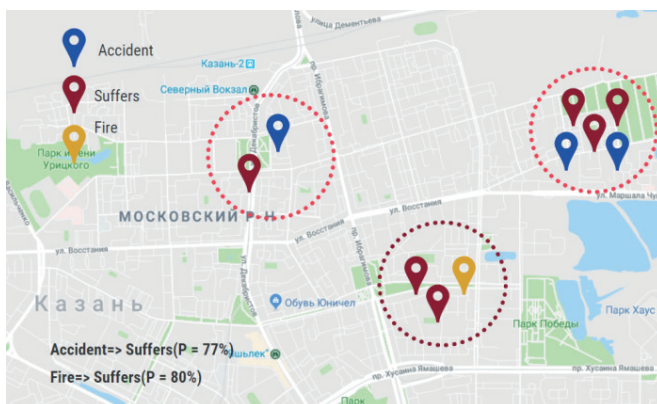


Рис. 3. Примеры полученных пространственных шаблонов
Fig. 3. Examples of obtained spatial patterns

2.2. Процесс выявления пространственных аномалий

Процесс выявления пространственных аномалий включает следующие шаги.

1. Пространственное деление. Во-первых, необходимо разделить пространственную область интереса на сетку равного размера, чтобы построить функцию временного ряда для каждой ячейки. Для этой задачи используется библиотека GeoSpark [12] для выполнения пространственного разбиения и выделения всех пространственных объектов, ограниченных в каждой полученной ячейке. GeoSpark предоставляет абстракцию данных SpatialRDD API для обработки пространственных данных и выполнения пространственных запросов в распределенной среде Spark. На рисунке 4 представлено пространственное деление на основе сетки;

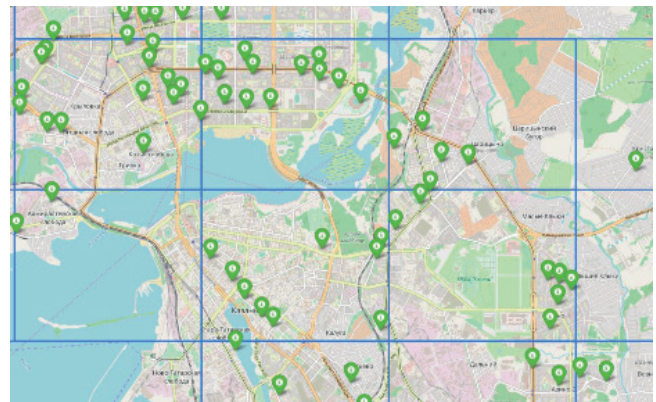


Рис. 4. Пространственное деление на основе сетки
Fig. 4. Spatial grid-based separation

2. Анализ временных рядов. Для каждой полученной пространственной ячейки выполняется анализ временных рядов. В рамках данной работы были применены модели ARIMA и рекуррентная нейронная сеть LSTM для прогнозирования временных рядов. Для модели ARIMA используется библиотека Python Statsmodels. Для обучения нейронной сети LSTM в распределенной среде используются библиотеки Keras[13] и Elephas[14]. Полученные модели и их графики представлены на рисунках 5 и 6.

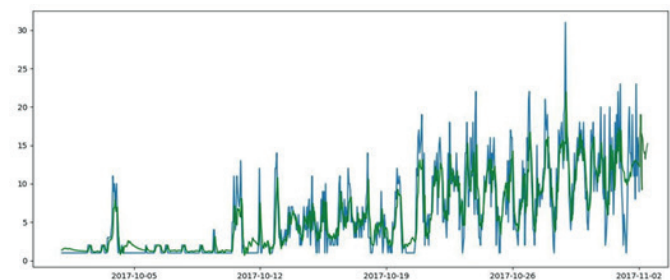


Рис. 5. Прогнозируемая модель ARIMA, X - момент времени,
Y - предсказанное количество событий
Fig. 5. Predicted model ARIMA, X - time, Y - predicted number of events



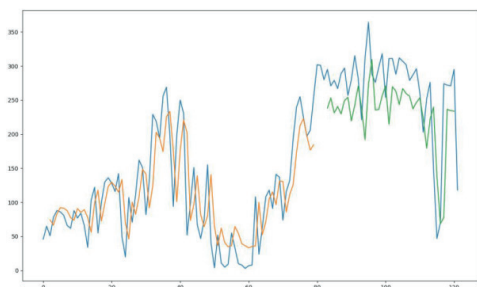


Рис. 6. Прогнозируемая модель LSTM, X - момент времени,
 Y - предсказанное количество событий

Fig. 6. Predicted model LSTM, X - point in time, Y - predicted number of events

2.3. Процесс анализа пространственной автокорреляции

GeoSpark используется для определения матрицы пространственных весов путем создания PolygonRDD и применения функции DistanceJoin для оценки расстояния между пространственными объектами. Оценка пространственной автокорреляции может осуществляться на двух уровнях: глобальном и локальном.

Для глобальных индексов Морана I, Geary's C и Getis - статистики Ord G используются для локального и LISA и Getis - Ord G* используется для статистических данных. Вычисление этой статистики производилось на платформе Apache Spark. Таким образом, пространственная автокорреляция была оценена для набора событий, представляющих собой дорожно-транспортные происшествия, произошедшие в вечернее время. Результаты представлены на рисунке 7.

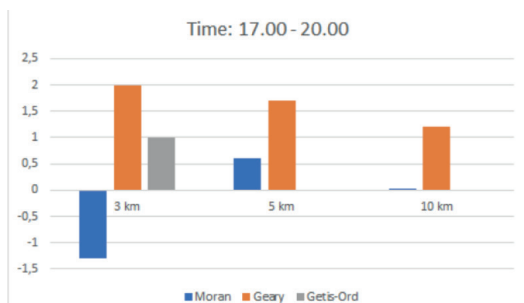


Рис. 7. Пространственная автокорреляция вечерних дорожно-транспортных происшествий

Fig. 7. Spatial autocorrelation of evening traffic accidents

3. Оценка производительности и эксперименты

Для проведения экспериментов в рамках разработанного программного комплекса в качестве входных данных был выбран набор данных, содержащий информацию о чрезвычайных ситуациях. Каждое событие в этом наборе данных характеризуется следующими атрибутами: геолокацией, временем возникновения события, типом события (преступления, вызов скорой, пожар и т. п.) и текстовым описанием события.

3.1. Экспериментальная установка

Вычислительный кластер на базе Apache Spark развертывается на пяти узлах со следующими характеристиками: 1) CPU: 2x6 ядер (2x HyperThreading) Intel Xeon CPU E5-2640, @2,50 ГГц; 2) Память (RAM): 128 ГБ.

3.2. Оценка эффективности

В разработанной системе реализованы следующие подходы: определения набора транзакций с помощью DBSCAN-кластеризации, интеллектуальный анализ ассоциативных правил с помощью FP-Growth, обнаружение аномалий с помощью моделей временных рядов и нейронных сетей, анализ пространственной автокорреляции. Поэтому оценка показателей эффективности для этих подходов проводилась отдельно для каждого этапа обработки. Эксперименты на рисунках 10-13 основаны на оценке временных характеристик в случае обработки различного объема данных в кластерной среде, развернутой на пяти машинах.

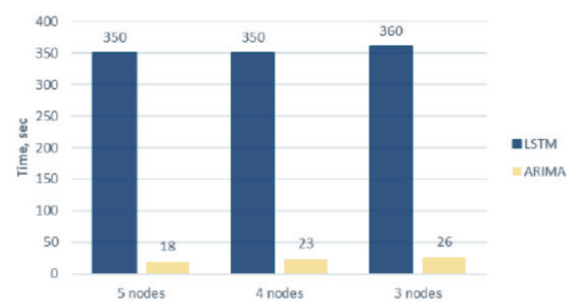


Рис. 8. Оценка времени выполнения для моделей временных рядов ARIMA и LSTM.

Fig. 8. Runtime estimation for time series models of ARIMA and LSTM

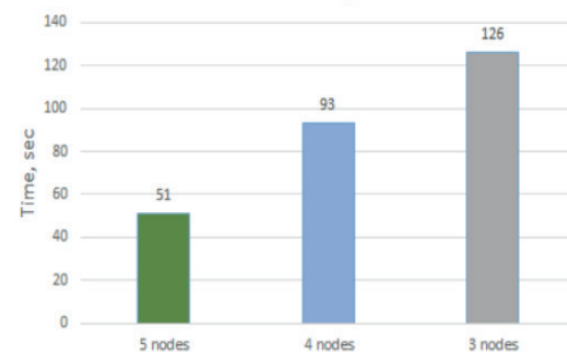


Рис. 9. Время выполнения кластеризации DBSCAN

Fig. 9. DBSCAN clustering runtime

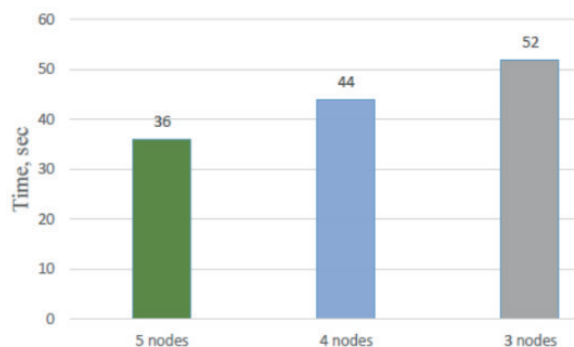
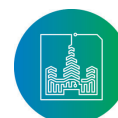


Рис. 10. Время выполнения интеллектуального анализа ассоциативных правил

Fig. 10. Mutual Rule Association Time



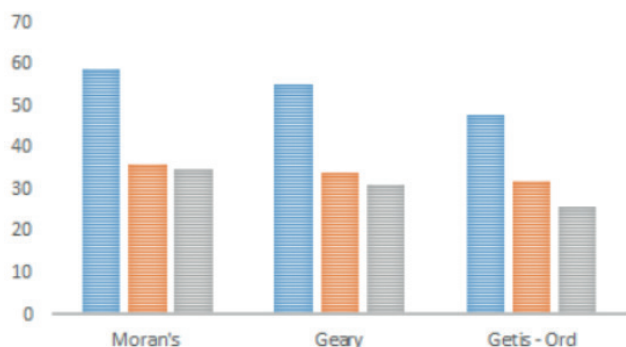


Рис. 11. Оценка времени для вычисления автокорреляции

Fig. 11. Estimated time to calculate autocorrelation

Заключение

В результате работы был реализован программный комплекс для анализа больших пространственно-временных данных. Все эксперименты проводились на разном объеме пространственно-временных данных, которые содержали информацию о чрезвычайных событиях. Рассмотренные подходы были реализованы на платформах распределенной обработки данных Apache Spark и Apache Hadoop. Рассмотренные подходы могут использоваться для анализа исторических данных, а также для обработки потоковых данных, поступающих в режиме реального времени с целью извлечения потенциально полезных знаний. Полученные знания могут быть использованы для ускорения процесса принятия решений при обработке чрезвычайных ситуаций. Кроме того, интеллектуальный анализ пространственных данных позволяет строить прогностические модели для предотвращения чрезвычайных ситуаций.

Список использованных источников

- [1] Roddick J.F., Spiliopoulou M. A bibliography of temporal, spatial and spatio-temporal data mining research // SIGKDD Explor. Newsl. 1999. Vol. 1, no. 1. Pp. 34-38. DOI: 10.1145/846170.846173
- [2] Shekhar S., Zhang P., Huang Y., Vatsavai R.R. Trends in spatial data mining / H. Kargupta, A. Joshi, K. Sivakumar, Y. Yesha (Ed.) // Data Mining: Next Generation Challenges and Future Directions. AAAI/MIT Press, 2004. Pp. 357-380.
- [3] Cressie N. Statistics for Spatial Data. Revised Edition. John Wiley & Sons, Inc., 1993. 900 p. DOI: 10.1002/9781119115151
- [4] Agrawal R., Srikant R. Fast Algorithms for Mining Association Rules // Proceedings of the 20th VLDB Conference Santiago, Chile, 1994. Pp. 487-499. URL: <http://www.vldb.org/conf/1994/P487.PDF> (дата обращения: 12.06.2018).
- [5] Huang Y., Zhang P. On the Relationships between Clustering and Spatial Co-location Pattern Mining // Proceedings of the 18th IEEE International Conference on Tools with Artificial Intelligence (ICTAI'06), Arlington, VA, 2006. Pp. 513-522. DOI: 10.1109/ICTAI.2006.91
- [6] Ester M., Kriegel H-P., Sander J., Xu X. A Density-Based Algorithm for Discovering Clusters a density-based algorithm for discovering clusters in large spatial databases with noise / E. Simoudis, J. Han, U. Fayyad (Eds.) // Proceedings of the

- Second International Conference on Knowledge Discovery and Data Mining (KDD'96). AAAI Press, 1996. Pp. 226-231.
- [7] Krishnamurthy R., Imielinski T. Practitioner problems in need of database research: Research directions in knowledge discovery // SIGMOD RECORD. 1991. Vol. 20, issue 3. Pp. 76-78.
- [8] Adhikari R., Agrawal R.K. An Introductory Study on Time series Modeling and Forecasting. LAP Lambert Academic Publishing, 2013. 69 p. DOI: 10.13140/2.1.2771.8084
- [9] Sak H., Senior A., Beaufays F. Long Short-Term Memory Recurrent Neural Network architectures for large scale acoustic modeling // arXiv:1402.1128v1. 2014. 5 pp.
- [10] Li Y., Cao H. Prediction for Tourism Flow based on LSTM Neural Network // Procedia Computer Science. 2018. Vol. 129. Pp. 277-283. DOI: 10.1016/j.procs.2018.03.076
- [11] Getis A. A History of the Concept of Spatial Autocorrelation: A Geographer's Perspective // Geographical Analysis. 2008. Vol. 40, issue 3. Pp. 297-309. DOI: 10.1111/j.1538-4632.2008.00727.x
- [12] Yu J., Wu J., Sarwat M. GeoSpark: A cluster computing framework for processing large-scale spatial data // Proceedings of the 23rd SIGSPATIAL International Conference on Advances in Geographic Information Systems (SIGSPATIAL'15). ACM, New York, NY, USA, 2015. Article 70, 4 pp. DOI: 10.1145/2820783.2820860
- [13] Keras library. URL: <https://keras.io> (дата обращения: 12.06.2018).
- [14] Elephas: Distributed Deep Learning with Keras & Spark. URL: <https://github.com/maxpumperla/elephas#usage-of-data-parallel-models> (дата обращения: 12.06.2018).
- [15] Garaeva A., Makhmutova E., Anikin I., Sattler K.-U. A framework for co-location patterns mining in big spatial data // Proceedings of the 20th IEEE International Conference on Soft Computing and Measurements (SCM 2017). St. Petersburg, 2017. Pp. 477-480. DOI: 10.1109/SCM.2017.7970622

Поступила 12.06.2018; принята в печать 10.09.2018;
опубликована онлайн 30.09.2018.

References

- [1] Roddick J.F., Spiliopoulou M. A bibliography of temporal, spatial and spatio-temporal data mining research. *SIGKDD Explor. Newsl.* 1999; 1(1):34-38. DOI: 10.1145/846170.846173
- [2] Shekhar S., Zhang P., Huang Y., Vatsavai R.R. Trends in spatial data mining. H. Kargupta, A. Joshi, K. Sivakumar, Y. Yesha (Eds.) *Data Mining: Next Generation Challenges and Future Directions*. AAAI/MIT Press, 2004, pp. 357-380.
- [3] Cressie N. *Statistics for Spatial Data*. Revised Edition. John Wiley & Sons, Inc., 1993. 900 p. DOI: 10.1002/9781119115151
- [4] Agrawal R., Srikant R. Fast Algorithms for Mining Association Rules. *Proceedings of the 20th VLDB Conference Santiago, Chile, 1994*, pp. 487-499. Available at: <http://www.vldb.org/conf/1994/P487.PDF> (accessed 12.06.2018).
- [5] Huang Y., Zhang P. On the Relationships between Clustering and Spatial Co-location Pattern Mining. *Proceedings of the 18th IEEE International Conference on Tools with Artificial Intelligence (ICTAI'06)*. Arlington, VA, 2006, pp. 513-522. DOI: 10.1109/ICTAI.2006.91



- [6] Ester M. Kriegel H-P., Sander J., Xu X. A Density-Based Algorithm for Discovering Clusters a density-based algorithm for discovering clusters in large spatial databases with noise. E. Simoudis, J. Han, U. Fayyad (Eds.) *Proceedings of the Second International Conference on Knowledge Discovery and Data Mining (KDD'96)*. AAAI Press, 1996, pp. 226-231.
- [7] Krishnamurthy R., Imielinski T. Practitioner problems in need of database research: Research directions in knowledge discovery. *SIGMOD RECORD*. 1991; 20(3):76-78.
- [8] Adhikari R., Agrawal R.K. An Introductory Study on Time series Modeling and Forecasting. LAP Lambert Academic Publishing, 2013. 69 p. DOI: 10.13140/2.1.2771.8084
- [9] Sak H., Senior A., Beaufays F. Long Short-Term Memory Recurrent Neural Network architectures for large scale acoustic modeling. *arXiv:1402.1128v1*. 2014. 5 pp.
- [10] Li Y., Cao H. Prediction for Tourism Flow based on LSTM Neural Network. *Procedia Computer Science*. 2018; 129:277-283. DOI: 10.1016/j.procs.2018.03.076
- [11] Getis A. A History of the Concept of Spatial Autocorrelation: A Geographer's Perspective. *Geographical Analysis*. 2008; 40(3):297-309. DOI: 10.1111/j.1538-4632.2008.00727.x
- [12] Yu J., Wu J., Sarwat M. GeoSpark: A cluster computing framework for processing large-scale spatial data. *Proceedings of the 23rd SIGSPATIAL International Conference on Advances in Geographic Information Systems (SIGSPATIAL'15)*. ACM, New York, NY, USA, 2015. Article 70, 4 pp. DOI: 10.1145/2820783.2820860
- [13] Keras library. Available at: <https://keras.io> (accessed 12.06.2018).
- [14] Elephas: Distributed Deep Learning with Keras & Spark. Available at: <https://github.com/maxpumperla/elephas#usage-of-data-parallel-models> (accessed 12.06.2018).
- [15] Garaeva A., Makhmutova F., Anikin I., Sattler K.-U. A framework for co-location patterns mining in big spatial data. *Proceedings of the 20th IEEE International Conference on Soft Computing and Measurements (SCM 2017)*. St. Petersburg, 2017, pp. 477-480. DOI: 10.1109/SCM.2017.7970622
- Submitted 12.06.2018; revised 10.09.2018; published online 30.09.2018.

About the authors:

Alina R. Garaeva, graduate student, Kazan National Research Technical University named after A. N. Tupolev – KAI (10 Karl Marx Str., Kazan 420111, Russia), ORCID: <http://orcid.org/0000-0002-3516-1424>, crazyalin91@mail.ru

Rifkat N. Minnikhanov, Doctor of Engineering Sciences, Professor, Corresponding Member of the Academy of Sciences of the Republic of Tatarstan, Director, State Budget Institution «Road Traffic Safety» (5 Orenburgskij trakt, Kazan 420000, Russia), ORCID: <http://orcid.org/0000-0001-9166-2955>, Priemnaya.Minnihanov@tatar.ru

Maria V. Dagaeva, Chief Specialist of the Center for Intelligent Transport Systems, State Budget Institution «Road Traffic Safety» (5 Orenburgskij trakt, Kazan 420000, Russia), ORCID: <http://orcid.org/0000-0002-5444-9669>

Sophya S. Kildeeva, master, Kazan National Research Technical University named after A. N. Tupolev – KAI (10 Karl Marx Str., Kazan 420111, Russia), ORCID: <http://orcid.org/0000-0001-8871-2632>

Igor V. Anikin, Doctor of Technical Sciences, Associate Professor, Kazan National Research Technical University named after A. N. Tupolev – KAI (10 Karl Marx Str., Kazan 420111, Russia), ORCID: <http://orcid.org/0000-0001-9478-4894>



This is an Open Access article distributed under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by/4.0>), which permits unrestricted reuse, distribution, and reproduction in any medium provided the original work is properly cited.

