

УДК 004.8

DOI: [10.26102/2310-6018/2023.40.1.026](https://doi.org/10.26102/2310-6018/2023.40.1.026)

## Подходы к обработке больших пространственно-временных данных ГЛОНАСС+112 в условиях неопределенности

И.В. Аникин✉, Г.Е. Петров

*Казанский национальный исследовательский технический университет им. А.Н. Туполева-КАИ, Казань, Российская Федерация  
anikinigor777@mail.ru✉*

**Резюме.** В работе исследуются подходы к обработке больших пространственно-временных данных в единой государственной информационной системе (ЕГИС) ГЛОНАСС+112 в условиях пространственной и временной неопределенности. Данная система предназначена для организации взаимодействия оперативных служб в Республике Татарстан (РТ), осуществления комплексного сбора и обработки данных, характеризующих различные инциденты, на основании звонков, поступивших на единый номер экстренных служб «112». Исследована производительность и масштабируемость различных операций по работе с большими данными в данной системе, адаптированных для использования в условиях неопределенности (запрос диапазона с порогом, JOIN, поиск  $k$ -ближайших соседей). Предложены новые подходы для решения задач формирования ассоциативных правил и кластеризации в условиях пространственной и временной неопределенности. Предложена модернизация алгоритма кластеризации пространственно-временных данных ST-DBSCAN. Данный алгоритм внедрен в схему формирования ассоциативных правил. Разработан программный комплекс формирования ассоциативных правил для пространственно-временных данных в условиях неопределенности. Программный комплекс осуществляет анализ не только данных ГЛОНАСС+112, но и информации о погоде, поступающей из внешних источников. Формируемые ассоциативные правила могут быть использованы для принятия решений и планирования ресурсов подразделениями различных оперативных служб. Это позволит повысить эффективность управления нежелательными инцидентами и чрезвычайными ситуациями.

**Ключевые слова:** интеллектуальный анализ данных, пространственно-временные данные, неопределенность, кластеризация, ассоциативные правила, управление чрезвычайными ситуациями.

**Для цитирования:** Аникин И.В., Петров Г.Е. Подходы к обработке больших пространственно-временных данных ГЛОНАСС+112 в условиях неопределенности. *Моделирование, оптимизация и информационные технологии*. 2023;11(1). URL: <https://moitvvt.ru/ru/journal/pdf?id=1307> DOI: 10.26102/2310-6018/2023.40.1.026

## Approaches to processing big spatiotemporal uncertain data in GLONASS+112

I.V. Anikin✉, G.E. Petrov

*Kazan National Research Technical University named after A.N. Tupolev-KAI, Kazan, Russian Federation  
anikinigor777@mail.ru✉*

**Abstract.** The paper examines some approaches to processing big spatiotemporal uncertain data in GLONASS+112. This system is used for managing interaction between operational services in the Republic of Tatarstan and collecting and processing data characterizing various incidents, based on calls received by a common emergency number "112". The performance and scalability of several basic operations for managing big data (query with threshold, JOIN,  $k$ -nearest neighbors algorithm) were

studied; they were adapted for operating data under spatial and temporal uncertainty. New approaches to clustering and associative rules mining for uncertain data are suggested. Modernization of ST-DBSCAN algorithm for clustering spatiotemporal data is proposed. This algorithm is integrated into the association rules mining process. The program complex for forming the associative rules for spatiotemporal data under uncertainty has been developed. The complex is applied to analyze GLONASS+112 data as well as the information about weather conditions obtained from external sources. The associative rules being formed can be used by various units in operating services for decision-making and resource-planning. This would help to increase the efficiency of managing the emergencies and undesired incidents.

**Keywords:** data mining, spatiotemporal data, uncertainty, clustering, associative rules, emergency management.

**For citation:** Anikin I.V., Petrov G.E. Approaches to processing big spatiotemporal uncertain data in GLONASS+112. *Modeling, Optimization and Information Technology*. 2023;11(1). URL: <https://moitvvt.ru/journal/pdf?id=1307> DOI: 10.26102/2310-6018/2023.40.1.026 (In Russ.).

## Введение

Единая государственная информационная система (ЕГИС) ГЛОНАСС+112 представляет собой систему взаимодействия оперативных служб в Республике Татарстан (РТ), позволяющую осуществлять отслеживание их ресурсов и автоматизированную обработку телефонных звонков от пострадавших. Применение ГЛОНАСС+112 нацелено на эффективное управление нежелательными инцидентами и чрезвычайными ситуациями [1].

В рамках ГЛОНАСС+112 осуществляется комплексный сбор и обработка данных, характеризующих различные инциденты, на основании звонков, поступивших на единый номер экстренных служб «112». Учетная карточка инцидента включает в себя следующую основную информацию: идентификатор события, его категория, текстовое описание, пространственная геолокация события (GPS координаты), время регистрации события [2].

Значительную актуальность при этом приобретает интеллектуальный анализ данных, накопленных по результатам зарегистрированных инцидентов, нацеленный на прогнозирование нежелательных событий, выявление корневых причин появления различных инцидентов, выявление скрытых взаимосвязей между ними и т. д. [3]. Для решения данной задачи могут применяться различные методы машинного обучения и обработки больших данных – кластерного анализа, поиска ассоциативных правил, прогнозирования временных рядов и т. д. [4]. Однако, применение данных методов может быть осложнено рядом обстоятельств, связанных со спецификой собираемых данных об инцидентах, в том числе:

– необходимостью учета как пространственной, так и временной составляющей, характеризующей появление инцидента. В данном случае необходимо вести речь об обработке пространственно-временных данных;

– наличием пространственной неопределенности, вызванной неточностью геолокации анализируемого события. Представленные в карточке инцидента GPS координаты характеризуют не точное место его возникновения, а привязку к базовой станции. В карточке инцидента представлена также область неопределенности, привязанная к базовой станции, внутри которой и регистрируется инцидент. Внутри области неопределенности можно только с некоторой долей вероятности утверждать о точном месте, в котором был зарегистрирован инцидент;

– необходимостью обработки больших массивов данных в приемлемые сроки. Для г. Казани количество звонков на номер «112» может достигать до нескольких десятков тысяч в сутки.

Целью данной работы является повышение эффективности интеллектуального анализа данных в ЕГИС ГЛОНАСС+112 за счет учета их специфик как больших пространственно-временных данных с пространственной неопределенностью.

Основными задачами, направленными на достижение цели, являются:

– выделение базовых операций по работе с большими данными, применяемыми в методах интеллектуального анализа данных;

– исследование производительности базовых операций, адаптированных для работы в условиях неопределенности;

– разработка новых подходов к интеллектуальному анализу больших пространственно-временных данных с пространственной неопределенностью. В данной работе предлагаются новые подходы для решения задач формирования ассоциативных правил и кластеризации.

### **Базовые операции по работе с пространственно-временными данными в условиях неопределенности и исследование их производительности**

Формализацию неопределенности пространственно-временных данных будем осуществлять путем указания их центров и радиусов неопределенности в пространстве и времени, а также плотностей вероятности появления событий в данных областях [5].

В качестве базовых операций при работе с такими данными будем рассматривать: запрос диапазона с указанием порога [6], оператор JOIN [7] и поиск k-ближайших соседей [8]. Адаптация указанных операций для возможной работы с неопределенными данными позволит применять классические схемы их обработки, в том числе в методах машинного обучения.

Пример операции запроса диапазона с порогом (Query with threshold) для неопределенной пространственной компоненты представлен на Рисунке 1. Даная операция предполагает выбор событий внутри прямоугольника (в общем виде многоугольника), имеющих время появления и геолокацию в указанном диапазоне. При этом осуществляется выбор событий, у которых вероятность нахождения внутри запроса превышает указанный порог. Для области неопределенности, связанной с появлением события А на Рисунке 1, данная вероятность равна 0.5, для области В – 0.15, для области С – 1.

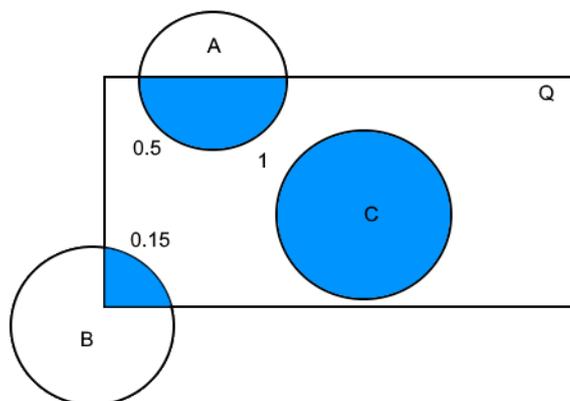


Рисунок 1 – Запрос с порогом в условиях неопределенности  
 Figure 1 – Query with threshold under uncertainty

Пример операции поиска k-ближайших соседей (kNN) для неопределенной пространственной компоненты представлен на Рисунке 2. При этом учитывается вероятность нахождения событий в области неопределенности. Вероятность события быть ближайшим соседом к некоторой точке определяется по формуле [9, 10]:

$$p_i = \int_0^{\infty} pr_i(r)dr * \prod_{k=1 \wedge k \neq i}^{|S|} (1 - P_k(r))$$

Здесь  $r$  – расстояние от центральной точки;  $P_k(r)$  – вероятность того, что искомое событие расположено ближе, чем  $r$  от центральной точки;  $pr_i(r)$  – плотность вероятности.

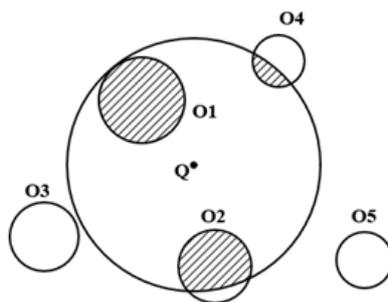


Рисунок 2 – Поиск k ближайших соседей в условиях неопределенности  
 Figure 2 – kNN query under uncertainty

Пример выполнения операции JOIN для неопределенной пространственной компоненты представлен на Рисунке 3 [11]. При этом решается задача нахождения вероятностей наличия объектов в области пересечения.

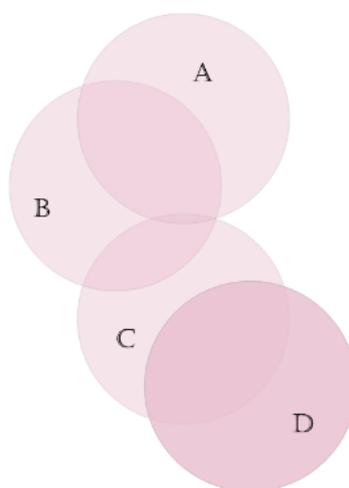


Рисунок 3 – Операция JOIN для пространственно-временных данных в условиях неопределенности  
 Figure 3 – JOIN operation under uncertainty

Наличие области неопределенности вместо конкретной координаты точки приводит к необходимости обработки большого количества данных при выполнении базовых операций. Это приводит к падению производительности.

Для повышения производительности базовых операций в условиях неопределенности можно использовать подходы, основанные на применении минимальных описывающих прямоугольников (Minimal Bounding Rectangle, MBR) [12] (Рисунок 4) и вероятностно ограничивающих прямоугольников (Probabilistically Constrained Rectangle, PCR) [13] (Рисунок 5), а также метод Монте-Карло.

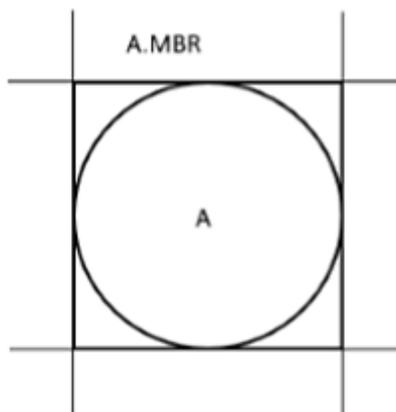


Рисунок 4 – Пространственный MBR для неопределенного объекта A  
 Figure 4 – Spatial MBR of an uncertain object A

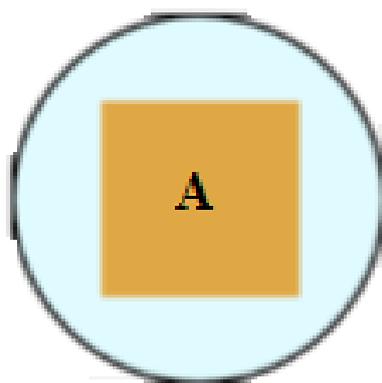


Рисунок 5 – Пространственный PCR для неопределенного объекта A  
 Figure 5 – Spatial PCR of an uncertain object A

Построение MBR осуществляется для каждой координаты точки, имеющей область неопределенности (например, для точек A, B, C, имеющих неопределенность в пространственной координате на Рисунке 1). Дальнейший анализ MBR вместо областей неопределенности позволяет значительно сократить количество анализируемых точек на предварительной стадии. Например, при выполнении операции запроса диапазона с порогом, на начальном этапе выполняется проверка наличия пересечения MBR с областью запроса. Если пересечение не подтверждается, то область неопределенности автоматически исключается из дальнейшего анализа, так как результирующая вероятность нахождения объекта в области запроса равна нулю. Если пересечение присутствует, то неопределенный объект обрабатывается обычным способом (Рисунок 1).

Область PCR применительно к области неопределенности представляет собой прямоугольник, полученный путем вдавливания внутрь каждой грани MBR до тех пор, пока вероятность появления неопределенного объекта внутри результирующей области не будет ниже заданного порога  $c$ . Области PCR имеют меньшую площадь по сравнению

с MBR, что уменьшает количество анализируемых точек. На Рисунке 6 представлен пример выполнения операции JOIN для 5 областей неопределенности с построением PCR областей.

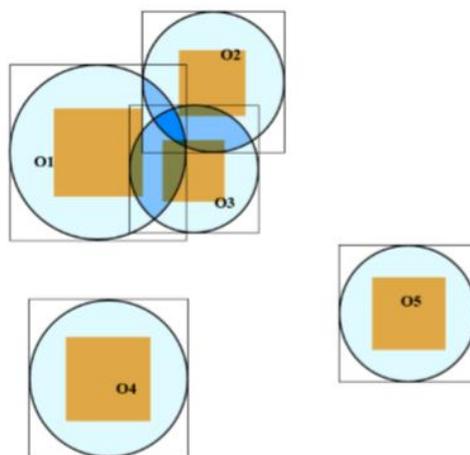


Рисунок 6 – PCR оптимизация для операции JOIN применительно к пространственно-неопределенным данным  
 Figure 6 – PCR optimization for uncertain spatial join

Применение метода Монте-Карло позволяет эффективно оценить вероятность попадания области неопределенности в формируемый результат базовой операции путем определения соответствующей площади фигуры. Рисунок 7 демонстрирует применение метода Монте-Карло для определения вероятности попадания пространственно-неопределенной точки A в запрос с порогом Q. Примеры подобных оценок вероятности представлены на Рисунке 1 (0.5 для области A, 0.15 для области B, 1 для области C).

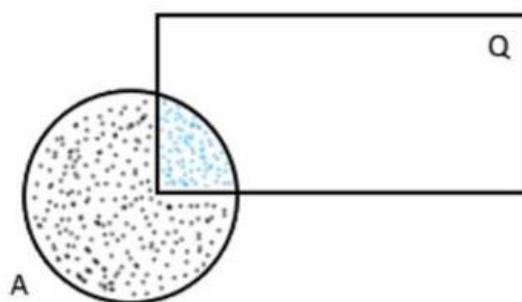


Рисунок 7 – Применение метода Монте-Карло для определения вероятности попадания пространственно-неопределенной точки A в запрос с порогом Q  
 Figure 7 – Monte-Carlo method for query with threshold under uncertain point A

Для исследования производительности и масштабируемости представленных расширений базовых операций в условиях неопределенности, авторами проведены соответствующие эксперименты. Реализация операций осуществлялась на платформе APACHE SPARK. Конфигурация вычислительного кластера, используемого для реализации разработанных операций включала в себя 3 узла, 36 ядер, 385 Мб ОЗУ. Для оценки производительности, разработанные операции тестировались на 3 наборах данных: 1.6GB, 6.2GB, 66.7GB, а также 2000, 5000, 10000 объектах. Полученные результаты тестирования представлены в Таблицах 2, 3.

Таблица 2 – Оценка производительности операций (на трех узлах)  
Table 2 – Evaluation of performance (for 3 nodes)

Операция	Время выполнения без оптимизации (сек)	Время выполнения с MBR оптимизацией (сек)	Время выполнения с PCR оптимизацией (сек)
Объем данных 1,6 Gb			
Запрос диапазона с порогом в условиях неопределенности	8926	38	33
Объем данных 6,2 Gb			
Запрос диапазона с порогом в условиях неопределенности	24021	98	82
Объем данных 66,7 Gb			
Запрос диапазона с порогом в условиях неопределенности	206602	768	758
2000 объектов			
Пространственно-временной оператор JOIN в условиях неопределенности	9098	155	142
5000 объектов			
Пространственно-временной оператор JOIN в условиях неопределенности	37882	979	750
10000 объектов			
Пространственно-временной оператор JOIN в условиях неопределенности	87880	3653	3255
	Время обработки данных 1.6 Gb (сек)	Время обработки данных 6.2 Gb (сек)	Время обработки данных 66.7 Gb (сек)
Нахождение k-ближайших соседей в условиях неопределенности	72	226	7356

Таблица 3 – Оценка масштабируемости операций (PCR оптимизация)  
Table 3 – Evaluation of scalability (PCR optimization)

Операция	Время выполнения операции с оптимизацией на одном узле (сек)	Время выполнения операции с оптимизацией на двух узлах (сек)	Время выполнения операции с оптимизацией на трех узлах (сек)
Диапазон с порогом	1117	964	758
к-ближайшие соседи	19042	11731	7356
JOIN	5021	3988	3255

Полученные результаты позволяют сделать вывод о том, что методы эффективной реализации выполнения базовых операций позволяют значительно снизить их вычислительную трудоемкость и обеспечить приемлемое время выполнения даже для значительных объемов данных.

### Формирование ассоциативных правил для пространственно-временных данных в условиях неопределенности

Ассоциативные правила представляют собой набор импликаций вида  $X \Rightarrow Y$ , где  $X \subset I$ ,  $Y \subset I$ ,  $X \cap Y = \emptyset$ , а  $I = \{i_1, \dots, i_n\}$  – набор некоторых элементов (событий), объединяемых в транзакции (типичные шаблоны)  $T$  [14]. Такие транзакции часто рассматриваются в качестве набора товаров, покупаемых покупателем за один визит. В случае событий, собираемых в ГЛОНАСС+112, в качестве транзакций могут рассматриваться совокупности событий, находящихся поблизости друг от друга во времени и в пространстве. В данном случае поиск ассоциативных правил позволит выявлять совокупности инцидентов, совместно встречаемых друг с другом примерно в одном месте и времени. При поиске таких правил следует учитывать пространственно-временную неопределенность данных.

Одним из подходов к учету неопределенности является формирование нечетких ассоциативных правил в виде:

ЕСЛИ  $X$  есть  $A$ , ТО  $Y$  есть  $B$

В данном случае  $X$  и  $Y$  рассматриваются как множества атрибутов, а  $A$ ,  $B$  – нечеткие множества, характеризующие  $X$  и  $Y$  соответственно. Возможные подходы к формированию таких правил были предложены в [15, 16].

В работе [17] авторами решалась задача построения ассоциативных правил с целью анализа эффекта влияния загрязнения окружающей среды на аллергию. Для формализации неопределенности, присущей пространственным атрибутам, а также рассматриваемым измерениям, использовались методы теории нечетких множеств [18].

В настоящей работе для учета пространственно-временной неопределенности при формировании ассоциативных правил предлагается особый способ формирования транзакций, основанный на модернизации алгоритма кластеризации пространственно-временных данных ST-DBSCAN [19]. При формировании кластеров анализируются две окрестности точки -  $\epsilon_1 = \text{SpatialEpsilon}$  (пространственная окрестность) и  $\epsilon_2 = \text{TimeEpsilon}$  (временная окрестность).

Примем равномерное распределение вероятностей появления событий в областях неопределенности пространства и времени. В этих условиях пространственные координаты анализируемых точек  $O_1, O_2, \dots$  будут выглядеть в виде, представленном на Рисунке 2.

Основой алгоритма ST-DBSCAN является подсчет количества точек, попадающих в пространственную и временную  $\varepsilon$ -окрестности некоторой точки  $A$ , и последующее сравнение полученной суммы точек с граничным значением  $\text{minPts}$ . Для работы ST-DBSCAN в условиях неопределенности введем следующую модернизацию – значения  $\varepsilon_1, \varepsilon_2$  будем откладывать от границы области неопределенности точки  $A$ . При этом возможны 3 ситуации (Рисунок 8):

а) анализируемая точка  $B$  полностью лежит внутри области неопределенности точки  $A + \varepsilon$ . В данном случае при подсчете количества точек, попавших в  $\varepsilon$ -окрестность  $A$ , точка  $B$  дает 100 % вклад к сумме  $\text{Neighbors}$ , сравниваемой с  $\text{minPts}$ . При суммарном подсчете точек должна добавляться единица (классический случай);

б) анализируемая точка  $B$  лежит частично внутри области неопределенности точки  $A + \varepsilon$ . В данном случае при подсчете количества точек, попавших в  $\varepsilon$ -окрестность  $A$ , точка  $B$  дает частичный (не 100 % вклад) к сумме  $\text{Neighbors}$ , сравниваемой с  $\text{minPts}$ . Данный вклад определяется площадью  $B$ , который можно рассчитать с помощью метода Монте-Карло. Приняв площадь области неопределенности точки  $B$  за 100 %, на этапе подсчета количества соседей  $\text{Neighbors}$ , к данной сумме должна прибавляться часть площади  $B$ , лежащая внутри области неопределенности точки  $A + \varepsilon$ . Данная площадь показана на Рисунке 9 (заштрихованная область);

в) анализируемая точка  $B$  лежит за пределами области неопределенности точки  $A + \varepsilon$ . В данном случае считаем, что она не достижима из точки  $A$ . При подсчете количества точек, попавших в  $\varepsilon$ -окрестность  $A$ , она дает 0 % вклад к сумме  $\text{Neighbors}$ , сравниваемой с  $\text{minPts}$ .

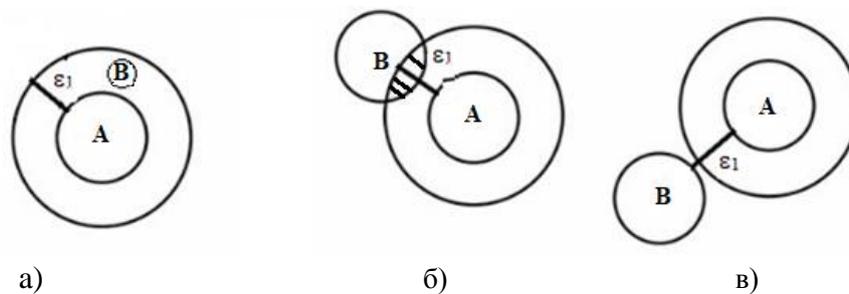


Рисунок 8 – Оценка расстояния между точками с учетом областей неопределенности и расстояния  $\varepsilon$

Figure 8 – Distance estimation between points accounting for uncertain regions and  $\varepsilon$  distance

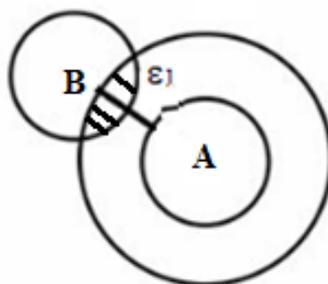


Рисунок 9 – Вклад точки  $B$  в сумму  $\text{Neighbors}$  (заштрихованная область)  
 Figure 9 – Contribution of point into the  $\text{Neighbors}$  sum (shaded area)

В данном случае, процедура RangeQuery в алгоритме ST-DBSCAN будет записана следующим образом:

```

RangeQuery(DB, A, ε)
{
  Neighbors=empty list
  for each point B in database DB {
    Neighbors = Square(A+ε,B)
  }
}
return Neighbors
}
    
```

В качестве итоговых транзакций при формировании ассоциативных правил будем рассматривать события, входящие в состав сформированных кластеров. Далее, для формирования ассоциативных правил будем использовать алгоритм FPGrowth [20]. В этом случае схема формирования ассоциативных правил для пространственно-временных данных в условиях неопределенности представлена на Рисунке 10.

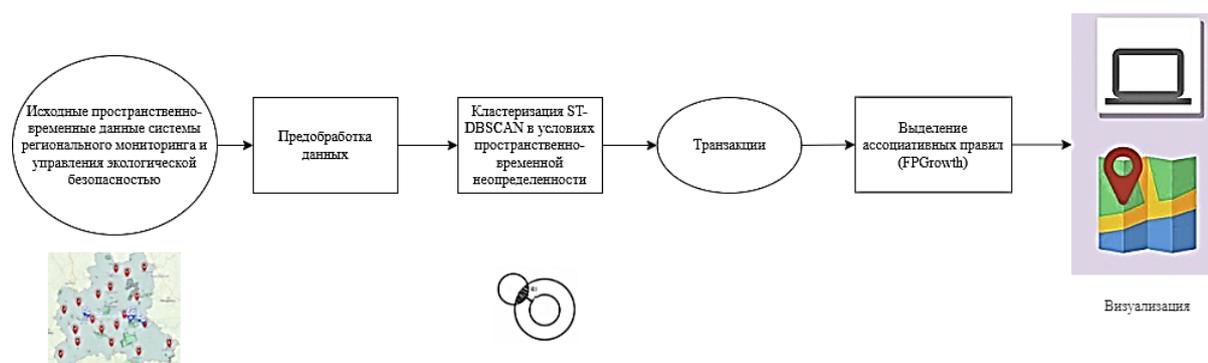


Рисунок 10 – Схема формирования ассоциативных правил для пространственно-временных данных в условиях неопределенности

Figure 10 – Method of associative rules mining for spatiotemporal data under uncertainty

### Программный комплекс поиска ассоциативных правил для пространственно-временных данных ГЛОНАСС+112 в условиях неопределенности

Авторами реализован на языке С# программный комплекс поиска ассоциативных правил для пространственно-временных данных ГЛОНАСС+112 в условиях неопределенности, реализующий схему, представленную на Рисунке 10. В программном комплексе реализованы алгоритм FPGrowth, модернизированный алгоритм ST-DBSCAN, алгоритм запроса диапазона с порогом. Для учета неопределенности при подсчете количества соседей Neighbors в модернизированной версии алгоритма ST-DBSCAN использован метод Монте-Карло. Дополнительно к основным категориям инцидентов ГЛОНАСС+112 (пожар, ДТП, криминал и другие – всего 14 категорий), осуществлялся анализ 28 погодных факторов, полученных из внешнего источника данных gr5.ru. Основное диалоговое окно программного комплекса представлено на Рисунке 11, на котором можно выделить пространственную область для дальнейшего анализа данных и отметить погодные факторы, включаемые или исключаемые из

дальнейшего анализа. Также возможно указание факторов, используемых только в условиях или при заключении ассоциативных правил.

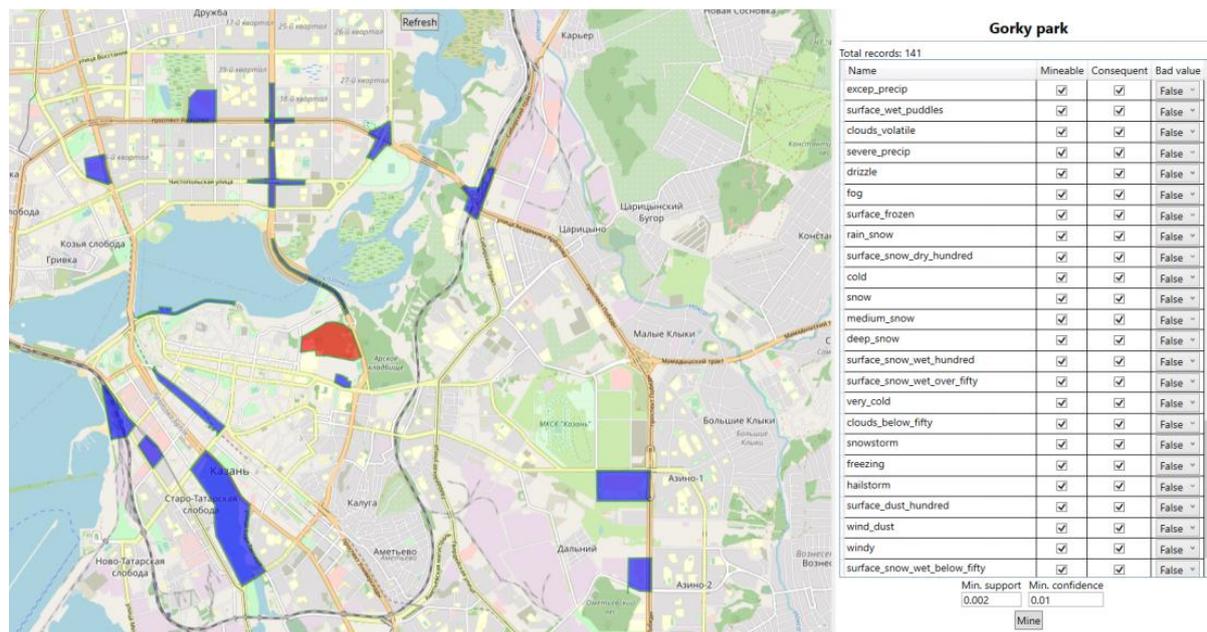


Рисунок 11 – Главное окно программного комплекса  
Figure 11 – Main window of the developed complex

Ниже представлены некоторые примеры сформированных правил, имеющие высокое значение показателя Support, с указанием области привязки (улица, объект и др.).

- {сильные осадки, снег, ДТП} -> вызов скорой помощи (Проспект Победы)
- {очень холодная температура, метель, ДТП} -> вызов скорой помощи (ул. Ленина)
- {глубокий снег, текущее выпадение снега} -> ДТП (ул. Ленина)
- {проблемы с дорожным трафиком, мокрая поверхность, ДТП} -> вызов скорой помощи (Проспект Победы)
- {проблемы с дорожным трафиком, очень холодная температура, ДТП} -> вызов скорой помощи (ул. 1 Мая)
- {высокая облачность, текущее выпадение снега, очень холодная температура} -> ДТП (ул. Амирхана/Чистопольская)
- {много снега на дороге, текущее выпадение снега, очень холодная температура} -> ДТП (ул. Проспект Победы)
- {беременность} -> вызов скорой помощи (ТЦ Тандем).

### Основные результаты работы и выводы

К основным результатам работы следует отнести следующие:

1. В работе выделены следующие базовые операции по работе с большими пространственно-временными данными, применяемыми в методах интеллектуального анализа данных: запрос диапазона с порогом, JOIN, поиск k-ближайших соседей. Исследован вопрос адаптации указанных базовых операций для работы в условиях неопределенности. Отмечено, что с целью повышения производительности их выполнения могут быть использованы подходы, основанные на применении MBR и PCR.

Для оценки вероятности попадания области неопределенности в формируемый результат базовой операции может быть использован метод Монте-Карло.

2. Исследована производительность базовых операций, адаптированных для работы в условиях неопределенности, при использовании MBR и PCR оптимизации, а также применении метода Монте-Карло для оценки вероятности попадания области неопределенности в формируемый результат базовой операции. Полученные результаты позволяют сделать вывод о том, что указанные методы оптимизации позволяют значительно снизить их вычислительную трудоемкость и обеспечить приемлемое время выполнения даже для значительных объемов данных. Это представляет актуальность для больших объемов неопределенных пространственно-временных данных ЕГИС ГЛОНАСС+112.

3. Предложены новые подходы для решения задач формирования ассоциативных правил и кластеризации в условиях пространственной и временной неопределенности. Предложена модернизация алгоритма кластеризации пространственно-временных данных ST-DBSCAN. Данный алгоритм внедрен в схему формирования ассоциативных правил. Для учета неопределенности при подсчете количества соседей в алгоритме кластеризации, использован метод Монте-Карло.

4. Разработан программный комплекс поиска ассоциативных правил для пространственно-временных данных ГЛОНАСС+112 в условиях неопределенности. Пространственная неопределенность данных вызвана неточностью их геолокации (привязкой к базовой станции). В разработанном программном комплексе реализованы алгоритм поиска ассоциативных правил FPGrowth, модернизированный алгоритм ST-DBSCAN, алгоритм запроса диапазона с порогом. Для учета неопределенности при подсчете количества соседей Neighbors в модернизированной версии алгоритма ST-DBSCAN, использован метод Монте-Карло.

Основные выводы:

1. Учет пространственной и временной неопределенности позволяет формировать более адекватную систему ассоциативных правил, ориентируясь не только на четко зафиксированное место и время появления события, но и принимая во внимание его область неопределенности (например, область привязки к базовой станции для ЕГИС ГЛОНАСС+112). Применение обычного подхода, заменяющего область неопределенности центральной точкой, делает систему ассоциативных правил менее адекватной к имеющимся реалиям.

2. Исследование производительности базовых операций запроса диапазона с порогом, JOIN, поиска k-ближайших соседей, адаптированных для работы в условиях пространственной и временной неопределенности, позволяет утверждать, что базовые операции имеют приемлемое время выполнения и могут быть использованы для решения более сложных задач машинного обучения, в частности, при решении задачи поиска ассоциативных правил. Для дальнейшего сокращения объема анализируемых данных, в диалоговом окне разработанного программного комплекса можно выделить на карте область для дальнейшего анализа данных и отметить погодные факторы, включаемые или исключаемые из дальнейшего анализа.

3. Сформированная система ассоциативных правил, осуществляющая связку анализируемых событий ЕГИС ГЛОНАСС+112, может быть использована для планирования ресурсов подразделениями различных оперативных служб и тем самым повысить эффективность управления нежелательными инцидентами и чрезвычайными ситуациями. Например, правило {проблемы с дорожным трафиком, мокрая поверхность, ДТП} -> вызов скорой помощи (Проспект Победы) говорит о необходимости принятия мер по искусственному ограничению скорости на соответствующем участке дороги.

## Заключение

В статье исследуются подходы к обработке больших пространственно-временных данных ГЛОНАСС+112 в условиях пространственной и временной неопределенности, что является неотъемлемым свойством данных, собираемых и обрабатываемых в данной системе. Исследована производительность и масштабируемость таких базовых операций работы с большими данными, как запрос диапазона с порогом, JOIN, поиск k-ближайших соседей, при использовании различных методов оптимизации, адаптированных для работы в условиях пространственной и временной неопределенности. Показано приемлемое время их выполнения, в том числе при решении задачи поиска ассоциативных правил.

Предложена модернизация алгоритма кластеризации пространственно-временных данных ST-DBSCAN для работы в условиях неопределенности. Данный алгоритм внедрен в схему формирования ассоциативных правил.

Разработан программный комплекс формирования ассоциативных правил для пространственно-временных данных в условиях неопределенности, осуществляющий анализ данных ГЛОНАСС+112 и информации из внешних источников (информации о погоде). В числе прочего, разработанный программный продукт предлагает выбор на карте областей для дальнейшего анализа (улицы, района и т. д.). Сформированная система ассоциативных правил, осуществляющая связку анализируемых событий, может быть использована для планирования ресурсов подразделениями различных оперативных служб и тем самым повысить эффективность управления нежелательными инцидентами и чрезвычайными ситуациями.

## СПИСОК ИСТОЧНИКОВ

1. Dagaeva M., Garaeva A., Anikin I., Makhmutova A., Minnikhanov R. Big spatio-temporal data mining for emergency management information systems. *IET Intelligent Transport Systems*. 2019;13(11):1649–1657.
2. Аникин И.В., Минниханов Р.Н., Дагаева М.В., Махмутова А.З., Чокоев А.Н. Программный комплекс выявления ассоциативных правил в ЕГИС ГЛОНАСС+112 с использованием внешних источников данных. *Сборник материалов форума KAZAN DIGITAL WEEK*. 2021:34–41.
3. Минниханов Р.Н., Дагаева М.В., Аникин И.В., Сабитов А.А., Гараева А.Р. Опыт применения технологий интеллектуального анализа данных в информационных системах Республики Татарстан. *Вестник ИЦБЖД*. 2021;2(48):159–167.
4. Aggarwal C.C. *Data mining. The textbook*. Springer Cham; 2014. 661 p.
5. Zheleznov A.N., Anikin I.V., Dagaeva M.V. Basic operations of analyzing uncertain spatio-temporal data and their application to data processing in GLONASS+112. *Proceedings of V International conference «Modern problems of life safety: intelligent transport systems and situation centers»*. 2018;2:240–245.
6. Thai M.T., Wu W., Xiong H. *Big Data in Complex and Social Networks*. Chapman & Hall. 2020:252.
7. Kromer P., Jurney R. *Big Data for Chimps: A Guide to Massive-Scale Data Processing in Practice*. O'Reilly Media. 2015:220.
8. Parsian M. *Data Algorithms. Recipes for scaling up with Hadoop and Spark*. O'Reilly Media, Inc. 2015:686.
9. Manocha S., Girolami M.A. An empirical analysis of the probabilistic K-nearest neighbour classifier. *Pattern Recognition Letters*. 2007;28(13):1818–1824.

10. Cheng R., Chen L., Chen J. Evaluating Probability Threshold k-Nearest-Neighbor Queries over Uncertain Data. *Proceedings of the 12th International Conference on Extending Database Technology: Advances in Database Technology*. 2009:672–683.
11. Lim H.-S., Lee J.-G., Lee M.-J., Whang K.-Y., Song I.-Y. Continuous query processing in data streams using duality of data and queries. *Proceedings of the ACM SIGMOD International Conference on Management of Data*. 2006:313–324.
12. Papadias D., Theodoridis Y. Spatial relations, minimum bounding rectangles, and spatial data structures. *International Journal of Geographical Information Science*. 1997;11(2):111–138.
13. Brisaboa N.R., Luaces M.R., Navarro G., Seco D. Range queries over a compact representation of minimum bounding rectangles. *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*. 2010;6413:33–42.
14. Agrawal R., Imieliński T., Swami A. Mining association rules between sets of items in large databases. *Proceedings of the 1993 ACM SIGMOD international conference on Management of data – SIGMOD '93*. 1993.
15. Kuok C.M., Fu A., Wong M.H. Mining Fuzzy Association Rules in Databases. *SIGMOD Record (ACM Special Interest Group on Management of Data)*. 1998;27(1):41–46.
16. Seda Unal Calargun, Adnan Yazici. Fuzzy association rule mining from spatio-temporal data. *Proceedings of Computational Science and Its Applications – ICCSA 2008 – International Conference*. 2008
17. Kanani Sadat Y., Nikaein T., Kar F. Fuzzy spatial association rule mining to analyze the effect of environmental variables on the risk of allergic asthma prevalence. *Geodesy and cartography*. 2015;41(2):101–112.
18. Zadeh L.A. Fuzzy Sets. *Information and Control*. 1965;8:338–363.
19. Birant D. ST-DBSCAN: An algorithm for clustering spatial-temporal data. *Data & Knowledge Engineering*. 2007;60(1):208–221.
20. Frequent Pattern Mining. Ed. Charu Aggarwal and Jiawei Han. *Springer*. 2014.

## REFERENCES

1. Dagaeva M., Garaeva A., Anikin I., Makhmutova A., Minnikhanov R. Big spatio-temporal data mining for emergency management information systems. *IET Intelligent Transport Systems*. 2019;13(11):1649–1657.
2. Anikin I.V., Minnikhanov R.N., Dagaeva M.V., Makhmutova A.Z., Chokoev A.N. The framework for associative rules mining in GLONASS+112 with using external source data. *Proceedings of Kazan Digital Week*. 2021:34–41. (In Russ.).
3. Minnikhanov R.N., Dagaeva M.V., Anikin I.V., Sabitov A.A., Garaeva A.R. Application of data mining techniques in information systems in Republic of Tatarstan. *Vestnik NCBJD = Proceedings of RCSO*. 2021;2(48):159–167. (In Russ.).
4. Aggarwal C.C. *Data mining. The textbook*. Springer Cham; 2014. 661 p.
5. Zheleznov A.N., Anikin I.V., Dagaeva M.V. Basic operations of analyzis uncertain spatio-temporal data and their application to data processing in GLONASS+112. *Proceedings of V International conference «Modern problems of life safety: intelligent transport systems and situation centers»*. 2018;2:240–245.
6. Thai M.T., Wu W., Xiong H. Big Data in Complex and Social Networks. *Chapman & Hall*. 2020:252.
7. Kromer P. *Journey R. Big Data for Chimps: A Guide to Massive-Scale Data Processing in Practice*. O'Reilly Media. 2015:220.

8. Parsian M. Data Algorithms. Recipes for scaling up with Hadoop and Spark. *O'Reilly Media, Inc.* 2015:686.
9. Manocha S., Girolami M.A. An empirical analysis of the probabilistic K-nearest neighbour classifier. *Pattern Recognition Letters*.2007;28(13):1818–1824.
10. Cheng R., Chen L., Chen J. Evaluating Probability Threshold k-Nearest-Neighbor Queries over Uncertain Data. *Proceedings of the 12th International Conference on Extending Database Technology: Advances in Database Technology*. 2009:672–683.
11. Lim H.-S., Lee J.-G., Lee M.-J., Whang K.-Y., Song I.-Y. Continuous query processing in data streams using duality of data and queries. *Proceedings of the ACM SIGMOD International Conference on Management of Data*. 2006:313–324.
12. Papadias D., Theodoridis Y. Spatial relations, minimum bounding rectangles, and spatial data structures. *International Journal of Geographical Information Science*. 1997;11(2):111–138.
13. Brisaboa N.R., Luaces M.R., Navarro G., Seco D. Range queries over a compact representation of minimum bounding rectangles. *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*. 2010;6413:33–42.
14. Agrawal R., Imieliński T., Swami A. Mining association rules between sets of items in large databases. *Proceedings of the 1993 ACM SIGMOD international conference on Management of data – SIGMOD '93*. 1993.
15. Kuok C.M., Fu A., Wong M.H. Mining Fuzzy Association Rules in Databases. *SIGMOD Record (ACM Special Interest Group on Management of Data)*. 1998;27(1):41–46.
16. Seda Unal Calargun, Adnan Yazici. Fuzzy association rule mining from spatio-temporal data. *Proceedings of Computational Science and Its Applications – ICCSA 2008 – International Conference*. 2008
17. Kanani Sadat Y., Nikaein T., Kar F. Fuzzy spatial association rule mining to analyze the effect of environmental variables on the risk of allergic asthma prevalence. *Geodesy and cartography*. 2015;41(2):101–112.
18. Zadeh L.A. Fuzzy Sets. *Information and Control*. 1965;8:338–363.
19. Birant D. ST-DBSCAN: An algorithm for clustering spatial-temporal data. *Data & Knowledge Engineering*. 2007;60(1):208–221.
20. Frequent Pattern Mining. Ed. Charu Aggarwal and Jiawei Han. *Springer*. 2014.

## ИНФОРМАЦИЯ ОБ АВТОРАХ / INFORMATION ABOUT THE AUTHORS

**Аникин Игорь Вячеславович**, доктор технических наук, профессор, заведующий кафедрой систем информационной безопасности, Казанский национальный исследовательский технический университет им. А.Н. Туполева-КАИ, Казань, Российская Федерация.

*e-mail:* [anikinigor777@mail.ru](mailto:anikinigor777@mail.ru)

ORCID: [0000-0001-9478-4894](https://orcid.org/0000-0001-9478-4894)

**Петров Глеб Евгеньевич**, аспирант кафедры систем информационной безопасности, Казанский национальный исследовательский технический университет им. А.Н. Туполева-КАИ, Казань, Российская Федерация.

*e-mail:* [gleb\\_petrov@mail.ru](mailto:gleb_petrov@mail.ru)

**Igor Vyacheslavovich Anikin**, Doctor of Technical Sciences, Professor, Head of Information Security Systems Department, Kazan National Research Technical University named after A.N. Tupolev-KAI, Kazan, Russian Federation.

**Gleb Evgenievich Petrov**, Postgraduate Student, Information Security Systems Department, Kazan National Research Technical University named after A.N. Tupolev-KAI, Kazan, Russian Federation.

*Статья поступила в редакцию 17.01.2023; одобрена после рецензирования 21.02.2023;  
принята к публикации 16.03.2023.*

*The article was submitted 17.01.2023; approved after reviewing 21.02.2023;  
accepted for publication 16.03.2023.*