

Для цитирования: Багутдинов Р.А. Подход к обработке, классификации и обнаружению новых классов и аномалий в разнородных и разномасштабных потоках данных. Вестник Дагестанского государственного технического университета. Технические науки. 2018;45 (3): 85-93. DOI:10.21822/2073-6185-2018-45-3-85-93

For citation: Bagutdinov R.A. Approach of processing, classification and detection of new classes and anomalies in heterogeneous and different streams of data. Herald of Daghestan State Technical University. Technical Sciences. 2018; 45(3):85-93. (In Russ.) DOI:10.21822/2073-6185-2018-45-3-85-93

ТЕХНИЧЕСКИЕ НАУКИ ИНФОРМАТИКА, ВЫЧИСЛИТЕЛЬНАЯ ТЕХНИКА И УПРАВЛЕНИЕ

УДК 004.04(004.6, 303.732)

DOI:10.21822/2073-6185-2018-45-3-85-93

ПОДХОД К ОБРАБОТКЕ, КЛАССИФИКАЦИИ И ОБНАРУЖЕНИЮ НОВЫХ КЛАССОВ И АНОМАЛИЙ В РАЗНОРОДНЫХ И РАЗНОМАСШТАБНЫХ ПОТОКАХ ДАННЫХ

Багутдинов Р.А.

Национальный исследовательский Томский политехнический университет,
634050, г.Томск, просп. Ленина, 30, Россия,
e-mail:lravil_bagutdinov@yahoo.com

Резюме. Цель. Целью исследования является поиск эффективных методов и подходов к обработке разнородных потоков данных и управления задачами бесконечной длины, концептуальной эволюции и концептуального дрейфа. Поток разнородных данных может иметь бесконечную длину и содержать структурированные или неструктурированные данные. Обработка разнородного и разномасштабного потока данных представляет собой основную проблему для исследователей. Большая часть исследований сосредоточена на решении проблемы бесконечной длины и концепции-дрейфа. **Метод.** Новые стратегии обнаружения класса классифицируются как параметрические и непараметрические. Данная работа основана на непараметрическом подходе. Классификатор работает на ансамбле трех моделей. Разделение генерирует различное количество классов в каждом фрагменте. Классы вычисляются путем применения метода К-медоидной кластеризации на каждом фрагменте. Эффективность метода К-медоидной кластеризации больше подходит для набора данных, содержащего аномалии. **Результат.** Разработанный алгоритм способен обрабатывать разнородные и разномасштабные данные. Каждый экземпляр, присутствующий в модели, принадлежит только одному классу. Экспериментальная работа была выполнена на четырех выборках потоковых данных по 2000 строк. После выполнения предварительной обработки в наборе данных были обнаружены многозначные характеристики данных. **Вывод.** В данной работе представлен эффективный подход для обработки разнородных потоков данных и управления задачами бесконечной длины, концептуальной эволюции и концептуального дрейфа. Разработанный подход основан на параметре сопоставления строк вместо расстояния для обработки четырех задач потоков данных. Уровень ложных срабатываний в разработанном алгоритме довольно низок и может считаться незначительным. Подход не классифицирует новый экземпляр класса как существующий класс, но может эффективно обрабатывать функциональную эволюцию.

Ключевые слова: поток данных, интеллектуальный анализ данных, разнородные данные, разномасштабные данные, обработка данных

TECHNICAL SCIENCE
COMPUTER SCIENCE, COMPUTER ENGINEERING AND MANAGEMENT

APPROACH OF PROCESSING, CLASSIFICATION AND DETECTION OF NEW CLASSES
AND ANOMALIES IN HETEROGENIOUS AND DIFFERENT STREAMS OF DATA

Ravil A. Bagutdinov

Tomsk Polytechnic University,
30 Lenina Ave., Tomsk 634050, Russia,
e-mail: lravil_bagutdinov@yahoo.com

Abstract. Objectives The aim of the study is to search for effective methods and approaches to the processing of heterogeneous data streams and the management of problems of infinite length, conceptual evolution and conceptual drift. A heterogeneous data stream can have infinite length and contain structured or unstructured data. Processing a heterogeneous and multi-scale data flow is a major challenge for researchers. Most of the research focuses on solving problems of infinite length and concept-drift. **Method.** New class detection strategies are classified as parametric and non-parametric. This work is based on a non-parametric approach. The classifier works on the ensemble of three models. The separation generates a different number of classes in each fragment. Classes are calculated by applying the K-Medoid clustering method on each fragment. The effectiveness of the K-media clustering method is more suitable for a data set containing anomalies. **Result.** The developed algorithm is capable of processing heterogeneous and multi-scale data. Each instance that is present in the model belongs to only one class. Experimental work was performed on four samples of stream data of 2000 lines each. After performing the pre-processing, the multi-valued characteristics of the data were found in the data set. **Conclusion.** This paper presents an effective approach for processing heterogeneous data streams and managing tasks of infinite length, conceptual evolution and conceptual drift. The developed approach is based on the string matching parameter instead of the distance for processing the four tasks of data streams. The level of false positives in the developed algorithm is rather low and can be considered insignificant. The approach does not classify a new instance of the class as an existing class, but can effectively handle the functional evolution.

Keywords: data flow, data mining, heterogeneous data, multiscale data, data processing

Введение. С развитием информационных технологий количество данных, генерируемых по каналам связи различных цифровых устройств, экспоненциально возрастают. Первоначально было сложно хранить и обрабатывать данные, генерируемые по каналу связи, но в настоящее время существует множество работ отечественных и зарубежных ученых, которые разработали методики и подходы для преодоления этого ограничения. Данные, сгенерированные в текстовом, аудио-, видеоформате и текущие от одного сетевого узла к другому, без прерывания, определяются как поток разнородных данных [1]. Основными характеристиками потоковых данных являются: непрерывность, динамический характер и неопределенный формат. Эти параметры постоянно меняются, что затрудняет процесс обработки. Ниже более подробно перечислены основные затруднения при обработке таких данных:

1. Поток данных генерируется с очень высокой скоростью и бесконечен по размеру. Сложно хранить и обрабатывать такой поток данных.
2. Предполагается, что дрейф присутствует, когда поток данных изменяется по времени. Это изменение заставляет исходным данным переходить (дрейфовать) от одних параметров и характеристик к другим.
3. Концептуальная эволюция возникает, когда новые данные эволюционируют в более новые данные. Например, эволюция происходит, когда обнаружен новый класс сигнатур вируса или обнаружен новый класс сетевой атаки. Такой эволюцией во время выполнения трудно управлять любой системой.

4. Функциональная эволюция - длительный процесс. Новая функция начинает появляться в потоке из-за концепции дрейфа и эволюции концепции. Эволюция новой функции влияет на существующие функции, и с течением времени в системе наблюдаются существенные изменения.

Статические методы классификации данных не могут использоваться при обработке потоков данных. Необходимо предложить эффективные методы и подходы классификации, которые подходят для решения задач потока данных [2,3]. Ниже приводятся основные проблемы и решения, существующие вместе с предлагаемым решением:

1. Проблема бесконечной длины: модели инкрементного обучения, в которых используется метод инкрементной обработки гибридной партии данных. Этот метод подразумевает разбиение потоков данных на кластеры одинаковых размеров для последующей их обработки.
2. Проблема дрейфа концепции: эту проблему можно выявить, отслеживая изменения, происходящие в потоковых данных. Изменения, происходящие в потоковых данных, являются переменными и обрабатываются моделями данных, которые требуют регулярных обновлений в соответствии с изменениями в потоковых данных. Здесь в основном применяется частичное решение. Изменчивость моделируется путем изменения количества классов данных, что позволяет системе обрабатывать новый класс данных.
3. Проблема эволюции концепции: в данной работе мы попытались решить эту проблему, позволяя классификатору автоматически определять новый класс без предварительной подготовки к новому классу.
4. Функциональная эволюция. Предлагаемое решение основано на операции сравнения строк. Сравниваются взаимодействие различных моделей, в которых обнаружены новые данные от каждой модели и выявляются аномалии в данных [4]. Вычисляется ошибка, которую используется для разделения экземпляров на основе их появления, т. е. концепции-эволюции, концепции дрейфа или шума. Новая функция используется для обновления существующей модели, чтобы она могла справляться с проблемами потоковой передачи данных.

Постановка задачи. Для решения задачи бесконечной длины и концепции дрейфа зарубежными учеными были предложены различные инкрементные подходы. В их числе: одноэтапный инкрементный подход и гибридный пакетный инкрементный подход. В инкрементном подходе для классификации используется только одна модель, которая динамически обновляется с регулярным временным интервалом. Гибридный инкрементный подход основан на совокупности разных моделей и методов пакетного обучения [5]. В этом подходе модель создается из последних данных и основана на эффективности классификации. Подход гибридной модели упрощает реализацию и обновление [6].

Выявление аномалий определяется как расстояние, измеренное между значением данных и всех других значений данных в случайной выборке данных. Проблема эволюции может быть решена путем выявления этих аномалий из данных.

Аномалии происходят в потоковых данных по таким причинам, как шум, эволюция концепции или дрейф концепции [7]. Необходимо выявить причины возникновения таких аномалий. Это позволит избежать ошибочной классификации концептуального дрейфа и уменьшить уровень ошибок (неправильно классифицировать существующий экземпляр класса как новый экземпляр класса) [8].

Новые стратегии обнаружения класса классифицируются как параметрические и непараметрические. Параметрический подход связывает нормальный диапазон данных с распределенным диапазоном для вычисления параметров распределения [9]. Если данные не соответствуют параметру распределения, они классифицируются как новый класс [10]. Непараметрические методы не основаны на распределении данных и, следовательно, не ограничены [11].

Методы исследования. Данная работа основана на непараметрическом подходе. Также большинство подходов, представленных в работах других авторов, могут обнаружить наличие только одного нового класса.

Предлагаемый автором подход. В данном случае, классификатор работает на ансамбле трех моделей. В предлагаемом подходе поток данных будет либо классифицирован в существующий класс, либо в новый класс. Пусть «L» представляет собой ансамбль моделей $\{M_1, M_2, M_3, \dots, M_n\}$.

В предлагаемом подходе используются следующие определения.

Определение 1. Существующий класс. Если модель M_i , принадлежащая ансамблю, обучается классу «С» и определяет ее, то класс «С» называется существующими классами. Другими словами, хотя бы одна модель, принадлежащая ансамблю M , должна обучаться классу С.

Определение 2. Новый класс. Если класс 'N' не известен ни одной из моделей M_i , принадлежащих ансамблю M , то «N» - новый класс. Ни одна модель ансамбля не была обучена новому классу.

Определение 3. Отклонения. Если x является тестовым экземпляром, и если он не соответствует спецификациям любого из классов С, то «x» является аномалией модели M_i . Аномалия не относится к какому-либо классу, определенному моделью.

На этапе обучения данные обучения делятся на равные кластеры. Для экспериментов и плавного управления размер каждого куска установлен на 2000 кортежей.

Разделение генерирует различное количество классов в каждом фрагменте [12]. Эти классы вычисляются путем применения метода К-медоидной кластеризации на каждом фрагменте. Эффективность метода К-медоидной кластеризации больше подходит для набора данных, содержащего аномалии [13].

На этапе обучения будет создана отдельная модель для каждого фрагмента данных, на которых проводится обучение [14]. Модель хранится как количество созданных кластеров и множеств (S_i), определяющих кластер. Правило классификации, за которым следует ансамбль: Если «x» - это экземпляр, подлежащий тестированию, он представляется каждой модели M_i в ансамбле, чтобы проверить, является ли это аномалией для модели M_i . Если он не является аномалией (Anom), он будет классифицирован по модели M_i в один из его классов, и если он будет обнаружен как аномалия всех трех моделей, тогда она будет считаться окончательной, то есть (FAnom).

На этапе обучения генерируются три модели, которые хранятся в виде количества кластеров и набора, определяющих кластеры [15]. Модели используются для обнаружения аномалий для тестовых данных. Данные испытания представляется каждой модели M_i для классификации.

Шаг 1. Чтобы классифицировать тестовый экземпляр данных, собираются кластеры, присутствующие в тестовом экземпляре данных [18]. Проверка выполняется, чтобы определить, присутствуют ли эти кластеры в наборе кластеров S_i , определяющих любой класс «С» модели.

Если кластеры тестового экземпляра присутствуют в множестве S_j класса 'C_j', то экземпляр классифицируется как принадлежащий классу 'C_j' модели 'M_i'. Если тестовый экземпляр данных не относится к какому-либо классу, определенному моделью «M_i», он объявляется как окончательный для этой модели «M_i».

Шаг 2. Этот шаг найдет окончательные аномалии ансамбля. Аномалия, обнаруженная на шаге 1 для каждой модели M_i , хранится в отдельном векторе «Anom_i». Каждый вектор «Anom_i» проверяется, чтобы узнать общий экземпляр данных, присутствующий во всех окончательных массивах. Если такой экземпляр найден, он объявляется как «FAnom», и все такие общие экземпляры хранятся в «FAnom_Vector». Процесс описан в алгоритме 1.

Алгоритм 1. FANOMVECTOR

Вход: Модели M_i и 'x'

Выход: FAnom_Vector (Вектор, содержащий аномалии в данных).

1. For each model 'M_i' in M
2. If $S(x) \in C_j M_i$ then
3. Append 'C_j' to 'x'
4. Else
5. Add 'x' to Anom_i
6. End if.

7. End for.

8. FAnom_Vector = Intersection (Anom_1, Anom_2.....Anom_i).

FAnom_Vector, сгенерированный в алгоритме1, содержит три типа аномалий. Они вызваны концептуальным дрейфом, эволюцией концепции и шумом.

Понятие-дрейф для экземпляра Anom_k из FAnom_Vector можно обрабатывать с помощью набора S_k и сравнения с множеством разных кластеров, принадлежащих модели M_i . Операция пересечения, выполняемая на множестве S_k , и множество S_j разных классов C_j , а если результат более 50%, он объявлен как аномалий из-за «концепции-дрейфа». Экземпляр этих данных сохраняется в векторе «CD».

Для обработки концепции-дрифта найден кластер или класс, к которому первоначально принадлежит экземпляр «Anom_k», и набор « S_j ».

Пусть « S_k » - это набор данных «Anom_k», присутствующего в векторе «CD». Разностная операция выполняется на двух наборах S_j и S_k . Набор результатов будет представлять собой набор новых кластеров и сохранен в векторе DW вместе с классом, из которой происходит дрейф. Для класса C_j модели M_i и уникальных кластеров построена матрица CM [m,j] со всеми элементами, установленными в 0. Матрица сканируется и для каждого нового дрейфа W_m класса C_j . Для каждого появления нового дрейфа W_m класса C_j значение в позиции CM [m, j] увеличивается на 1.

Для каждого уникального кластера установлено общее пороговое значение для сравнения. Если значение в CM [m, j] больше заданного порогового значения, тогда W_m добавляется к набору S_j класса C_j . Происходит обработка концепции дрейфа и правильно сдвигает новые кластеры в доступных классах. Процесс объясняется в Алгоритме 2 ниже.

Алгоритм 2. CONCEPTDRIFT

Вход: FVector and Model 'Mi'

Выход: CD (данные с концептуальным дрейфом и обновленной моделью).

1. For each Anom_k in FAnom_Vector
2. For each cluser C_j in model M_i
3. Result ← Set-Intersection (S_k, S_j)
4. If ($\neg(\text{Anom}(C_j))$ and ($S(\text{Result}) \geq (50\% \text{ of } C_j)$)) then
5. CD ← Anom_k
6. Store information about C_j in JCount.
7. End if
8. End for.
9. End for.
10. For each instance 'x' in CD belonging to cluster C_j
11. DW ← Set-Difference (X_i, S_j)
12. End for.
13. Unique_driftword ← Unique (DW).
14. For each ' W_m ' in Unique_DW
15. For each Class C_j in M_i
16. CM [m,j] ← CM [m, j]+1.
17. End for
18. End for
19. If (CM [m,j] > Threshold) then
20. Append word W_m ' S_j ' of Class ' C_j '
21. End if.
22. End algorithm.

Эволюция концепции определяется путем рассмотрения FAnom_Vector. Если экземпляр Anom_k, не удовлетворяющий критериям дрейфа концепции, он объявлен как эволюция концепции и сохраняется в векторе SE.

Экземпляр классифицируется как эволюция концепции, если более 50% кластеров экземпляра данных не удовлетворяют условию дрейфа концепции алгоритма 2.

Пороговое значение 50% фиксировано на основе проведенных экспериментов. В проведенных экспериментах отмечается, что порог 50% подходит для объявления экземпляра как эволюции концепции.

Эволюция концепции обрабатывается путем создания нового класса на основе результатов [16].

Процесс обработки концепции эволюции также связан с созданием нового класса [17]. Чтобы создать новый класс или новый класс, алгоритм кластеризации применяется к вектору CE. Число кластеров равно числу классов в векторе CE. Кластеры присоединяются к модели M_i ансамбля.

Алгоритм 3. CONEVOLUTION.

Вход: FVector and Model ' M_i '

Выход: CE (вектор, имеющий экземпляры concept_evolution)

1. For each Anom_k in FAnom_Vector.
2. For each cluster C_j in model M_i
3. Result ← Set Intersection (S_k, S_j)
4. If ((Anom (C_j)) and ($S(\text{Result}) < (50\% \text{ of } C_j)$)) then
5. CE ← Anom_k.
6. End if
7. End for
8. End for
9. Применим кластеризацию K-modoid на CE.
10. Получаем новые кластеры.
11. Добавляем новые кластеры в любые предыдущие модели M_i .
12. End algorithm

Обсуждение результатов. Разработанный алгоритм способен обрабатывать разнородные и разномасштабные данные. Каждый экземпляр, присутствующий в модели, принадлежит только одному классу. Экспериментальная работа была выполнена на четырех выборках потоковых данных по 2000 строк. После выполнения предварительной обработки в наборе данных были обнаружены многозначные характеристики данных.

Каждое событие имеет связанную с ним аномалию и рассматривается как другой класс [18]. Данные также содержат различные атрибуты, строки и столбцы, содержащие неполную информацию [19]. На этапе предварительной обработки удалены все такие строки и столбцы из набора данных. Набор данных содержит шесть нормальных классов и два новых класса.

В наборе данных, выбранном для экспериментов, ни один экземпляр, принадлежащий новому классу, не был объявлен как существующий экземпляр класса в наборе данных, и очень немногие экземпляры, принадлежащие к существующему классу, были объявлены как новый экземпляр класса [20]. Набор данных также содержит шум в виде экземпляров, принадлежащих к существующему классу, но остался неклассифицированным.

В табл. 1 показаны ошибки модели. Ошибка определяется как процент не классифицированных аномалий в данных. Уровень ложных срабатываний представляет собой процент ошибочно классифицированных случаев.

Из табл. 1 видно, что с каждым новым прогоном алгоритма процент ошибок снижается в среднем по частоте ошибок на 11,7, ошибок классификации на 0,4, уровень ложных срабатываний на 0,6%.

Таблица 1. Результаты ошибок при проверке подхода
Table 1. The results of errors in the verification approach

№ замера Measuring	Частота ошибок Error Frequency	Ошибочная классификация Erroneous classification	Уровень ложных срабатываний (%) The level of false positives (%)
1	14,65	1,2	0,4
2	2,55	0,6	1
3	2,1	0,8	-

Вывод. Представлен новый подход обработки разнородных и разномасштабных потоков данных. Представленный подход может обрабатывать бесконечную длину, концептуальную эволюцию и концептуальный дрейф. Он также может обнаруживать несколько новых классов, происходящих одновременно.

Разработанный подход основан на параметре сопоставления строк вместо расстояния для обработки четырех задач потоков данных. Уровень ложных срабатываний в разработанном алгоритме довольно низок и может считаться незначительным. Подход не классифицирует новый экземпляр класса как существующий класс, но может эффективно обрабатывать функциональную эволюцию. Все эксперименты проводились на данных фиксированного размера. В будущем планируется проверка подхода на динамических потоках данных.

Библиографический список:

1. A. El-Hoiydi, "Spatial TDMA and CSMA with preamble sampling for low power ad hoc wireless sensor networks", Proceedings of ISCC 2002, Seventh International Symposium on Computers and Communications, pp. 685 - 692, July 2002.
2. Bagutdinov R.A. The processing of heterogeneous data for multisensor systems of technical vision on the example of analysis of temperature and gas concentration / MSIT TPU. 2018. P. 25-26.
3. Багутдинов Р.А. Принцип разработки алгоритмического обеспечения системы технического зрения роботов / Научные технологии в космических исследованиях Земли. 2017. Т. 9. №5. С. 66-71.
4. Bagutdinov R.A., Zaharova A.A. The task adaptation method for determining the optical flow problem of interactive objects recognition in real time / Journal of Physics: Conference Series. 2017. Т. 803. №1. С. 012014.
5. Барсегян А.А. Технологии анализа данных. Data Mining, Visual Mining, Text Mining, OLAP / Барсегян А.А., Куприянов М.С., Степаненко В.В., Холод И.И // СПб.: БХВ-Петербург, 2007.- 384 с.: ил.
6. Ganguly AR, Steinhäuser K Data mining for climate change and impacts. In Proceedings of IEEE international conference on data mining (ICDM) workshops. 2009; pp 385–394
7. Кадиев И.П., Кадиев П.А. Основы индексной структуризации nxn - комбинаторных конфигураций // Вестник Дагестанского государственного технического университета. Технические науки. 2018;45(1):139-146
8. Качаева Г.И., Попов А.Д., Рогозин Е.А. Показатели эффективности функционирования при разработке систем защиты информации от несанкционированного доступа в автоматизированных информационных системах // Вестник Дагестанского государственного технического университета. Технические науки. 2018;45(1):147-159.
9. Kuei-PingShih , Hung-ChangChen , Chien-MinChou , Bo-JunLiu "On target coverage in wireless heterogeneous sensor networks with multiple sensing units", Journal of Network and Computer Applications, vol. 32, pp. 866– 877, 2009.
10. Островский О.А. Дефиниционный анализ корреляционной зависимости информационной модели и криминалистической характеристики преступления в сфере компьютерной информации / Евразийский юридический журнал. №7 (110). 2017. С. 221-225.
11. Островский О.А. Принцип объектной декомпозиции в систематизации идентификационных кодов, характеризующих преступления в сфере компьютерной информации / Полицейская деятельность. № 3. 2017. С. 10-18.
12. Островский О.А. Криминалистический анализ, описывающий состояние детерминированного конечного автомата в модели наблюдателя при расследовании преступлений в сфере компьютерной информации / Евразийский юридический журнал. №3 (118). 2018. С. 294-296.
13. Островский О.А. Алгоритмы проведения осмотров цифровых носителей информации для предотвращения компьютерных преступлений / Военно-юридический журнал. № 11. 2017. С. 3-6.
14. Pietro Ciciriello, Luca Mottola, Gian Pietro Picco, "Efficient routing from multiple sources to multiple sinks in wireless sensor networks", in Proceedings of the 4th European Conference on Wireless Sensor Networks (EWSN'07), Lecture Notes in Computer Science, vol. 4373, pp. 34–50, January 2007.
15. Петренко Н.А., Багутдинов Р.А. Анализ мультисенсорных систем и сенсорного слияния данных / В сборнике: Молодёжь и современные информационные технологии. Сборник трудов XV Международной научно-практической конференции студентов, аспирантов и молодых учёных. Национальный исследовательский Томский политехнический университет. 2018. С. 73-74.
16. Ramaswamy S, Rastogi R, Shim K Efficient algorithms for mining outliers from large data sets. ACM SIGMOD. 2000; Rec 29(2):427–438
17. Hart JK, Martiez K Environmental sensor networks: a revolution in the earth system sciene? 2006; Earth Sci Rev 78:177–191
18. Юркова О.Н. Применение методов анализа данных для автоматизации формирования онтологии. // Вестник Дагестанского государственного технического университета. Технические науки. 2018;45(1):172-180.
19. W. Ye, J. Heidemann, and D. Estrin, "Medium Access Control with Coordinated Adaptive Sleeping for Wireless Sensor Networks", IEEE/ACM Trans. Networking, 2004; vol. 12, no. 3, pp. 493–506

20. Xie M, Hu J, Tian B Histogram-based online anomaly detection in hierarchical wireless sensor networks. In: Trust, Security and Privacy in Computing and Communications, 2012 IEEE 11th International Conference On. IEEE. 751–759

References:

1. A. El-Hoiydi, "Spatial TDMA and CSMA with preamble sampling for low power ad hoc wireless sensor networks", Proceedings of ISCC 2002, Seventh International Symposium on Computers and Communications, pp. 685 - 692, July 2002.
2. Bagutdinov R.A. The processing of heterogeneous data for multisensor systems of technical vision on the example of analysis of temperature and gas concentration / MSIT TPU. 2018. P. 25-26.
3. Bagutdinov R.A. Printsip razrabotki algoritmicheskogo obespecheniya sistemy tekhnicheskogo zre-niya robotov / Naukoyemkiye tekhnologii v kosmicheskikh issledovaniyakh Zemli. 2017. T. 9.№5. S. 66-71. [Bagutdinov R.A. The principle of developing algorithmic support of the robotic vision system / High technology in space exploration of the Earth.2017. T. 9.№5. P. 66-71. (In Russ.)]
4. Bagutdinov R.A., Zaharova A.A. The task adaptation method for determining the optical flow problem of interactive objects recognition in real time / Journal of Physics: Conference Series. 2017. T. 803. №1. C. 012014.
5. Barsegyan A.A. Tekhnologii analiza dannykh. Data Mining, Visual Mining, Text Mining, OLAP / Barsegyan A.A., Kupriyanov M.S., Stepanenko V.V., Kholod I.I // SPb.: BKHV-Peterburg, 2007.- 384 s.: il. [Barsegyan A.A. Data Analysis Technologies. Data Mining, Visual Mining, Text Mining, OLAP / Barsegyan AA, Kupriyanov MS, Stepanenko VV, Kholod II // St. Petersburg: BHV-Petersburg, 2007.- 384 p. (In Russ.)]
6. Ganguly AR, Steinhäuser K Data mining for climate change and impacts. In Proceedings of IEEE international conference on data mining (ICDM)workshops. 2009; pp 385–394
7. Kadiyev I.P., Kadiyev P.A. Osnovy indeksnoy strukturizatsii nxn - kombinatorykh konfiguratsiy./Vestnik Dagestanskogo gosudarstvennogo tekhnicheskogo universiteta. Tekhnicheskkiye nauki. 2018;45(1):139-146 [Kadiyev IP, Kadiyev PA Fundamentals of index structuring of nxn - combinatorial configurations. / Herald of Dagestan State Technical University. Technical science. 2018;45(1):139-146. (In Russ.)]
8. Kachayeva G.I., Popov A.D., Rogozin Ye.A. Pokazateli effektivnosti funktsionirovaniya pri razrabotke sistem zashchity informatsii ot nesanksionirovannogo dostupa v avtomatizirovannykh informatsionnykh sistemakh./Vestnik Dagestanskogo gosudarstvennogo tekhnicheskogo universiteta. Tekhnicheskkiye nauki. 2018;45(1):147-159. [Kachayeva G.I, Popov A.D, Rogozin E.A. Performance indicators for the development of information security systems against unauthorized access in automated information systems. /Herald of the Dagestan State Technical University. Technical science. 2018;45(1):147-159. (In Russ.)]
9. Kuei-Ping Shih , Hung-Chang Chen , Chien-MinChou , Bo-JunLiu "On target coverage in wireless heterogeneous sensor networks with multiple sensing units", Journal of Network and Computer Applications, vol. 32, pp. 866– 877, 2009.
10. Ostrovskiy O.A. Definitivnyy analiz korrelyatsionnoy zavisimosti informatsionnoy modeli i kriminalisticheskoy kharakteristiki prestupleniya v sfere komp'yuternoy informatsii / Yevraziyskiy yuridicheskiy zhurnal. №7 (110). 2017. S. 221-225. [Ostrovsky OA A definitive analysis of the correlation dependence of the information model and the forensic characteristics of the crime in the sphere of computer information / Eurasian Juridical Journal. №7 (110). 2017. P. 221-225. (In Russ.)]
11. Ostrovskiy O.A. Printsip ob'yektnoy dekompozitsii v sistematizatsii identifikatsionnykh kodov, kharakterizuyushchikh prestupleniya v sfere komp'yuternoy informatsii / Politseyskaya deyatel'nost'. № 3. 2017. S. 10-18. [Ostrovsky OA The principle of object decomposition in the systematization of identification codes characterizing crimes in the sphere of computer information / Police activity.№3. 2017. P. 10-18. (In Russ.)]
12. Ostrovskiy O.A.Kriminalisticheskyy analiz, opisyyvayushchiy sostoyaniye determinirovannogo ko-nechnogo avtomata v modeli nablyudatelya pri rassledovanii prestupleniy v sfere komp'yuternoy informatsii / Yevraziyskiy yuridicheskiy zhurnal. №3 (118). 2018. S. 294-296. [Ostrovsky OA Criminalistic analysis describing the state of a deterministic finite automaton in the observer model in the investigation of crimes in the sphere of computer information / Eurasian Juridical Journal. №3 (118). 2018. C. 294-296. (In Russ.)]
13. Ostrovskiy O.A.Algoritmy provedeniya osmotrov tsifrovyykh nositeley informatsii dlya predotvrashcheniya komp'yuternyykh prestupleniy / Voyenno-yuridicheskiy zhurnal. № 11. 2017. S. 3-6. [Ostrovsky OA Algorithms for carrying out examinations of digital media for preventing computer crimes / Military-legal journal. № 11. 2017. pp. 3-6. (In Russ.)]
14. Pietro Ciciriello, Luca Mottola, Gian Pietro Picco, "Efficient routing from multiple sources to multiple sinks in wireless sensor networks", in Proceedings of the 4th European Conference on Wireless Sensor Networks (EWSN'07), Lecture Notes in Computer Science, vol. 4373, pp. 34–50, January 2007.
15. Petrenko N.A., Bagutdinov R.A. Analiz mul'tisensornykh sistem i sensorного sliyaniya dannykh / V sbornike: Molodozh' i sovremennyye informatsionnyye tekhnologii. Sbornik trudov XV Mezhdunarodnoy nauchno-prakticheskoy konferentsii studentov, aspirantov i molodykh uchonykh. Natsional'nyy issledovatel'skiy Tomskiy politekhnicheskii universitet. 2018. S. 73-74. [Petrenko NA, Bagutdinov RA Analysis of multi-sensory systems and sensory data merging / In the collection: Youth and modern information technology. Proceedings of the XV International Scientific and Practical Conference of students, graduate students and young scientists. National Research Tomsk Polytechnic University. 2018; P. 73-74. (In Russ.)]

16. Ramaswamy S, Rastogi R, Shim K Efficient algorithms for mining outliers from large data sets. ACM SIGMOD. 2000; Rec 29(2):427–438
17. Hart JK, Martiez K Environmental sensor networks: a revolution in the earth system sciene? 2006; Earth Sci Rev 78:177–191.
18. Yurkova O.N. Primeneniye metodov analiza dannykh dlya avtomatizatsii formirovaniya ontologii. / Vestnik Dagestanskogo gosudarstvennogo tekhnicheskogo universiteta. Tekhnicheskiye nauki. 2018;45(1):172-180.[Yurkova ON Application of data analysis methods to automate the formation of ontology. / Herald of the Dagestan State Technical University. Technical science. 2018;45(1):172-180. (In Russ.)]
19. W. Ye, J. Heidemann, and D. Estrin, “Medium Access Control with Coordinated Adaptive Sleeping for Wireless Sensor Networks”, IEEE/ACM Trans. Networking, 2004;vol. 12, no. 3, pp. 493–506.
20. Xie M, Hu J, Tian B Histogram-based online anomaly detection in hierarchical wireless sensor networks. In: Trust, Security and Privacy in Computing and Communications, 2012 IEEE 11th International Conference On. IEEE. 751–759

Сведения об авторе:

Багуддинов Равиль Анатольевич – ассистент, отделение автоматизации и робототехники Инженерной школы информационных технологий и робототехники.

Information about the author.

Ravil A. Bagutdinov– Assistant, Department of Automation and Robotics, School of Information Technology and Robotics.

Конфликт интересов.

Автор заявляет об отсутствии конфликта интересов.

Поступила в редакцию 28.06.2018.

Принята в печать 29.08.2018.

Conflict of interest.

The author declare no conflict of interest.

Received 28.06.2018.

Accepted for publication 29.08.2018.