



<http://dx.doi.org/10.35596/1729-7648-2019-124-6-62-71>

Оригинальная статья
Original paper

УДК 519.684.6:004.021

ИНТЕЛЛЕКТУАЛЬНЫЙ АНАЛИЗ ДАННЫХ И ОБЛАЧНЫЕ ВЫЧИСЛЕНИЯ

ТАТУР М.М., ЛУКАШЕВИЧ М.М., ПЕРЦЕВ Д.Ю., ИСКРА Н.А.

Белорусский государственный университет информатики и радиоэлектроники, Республика Беларусь

Поступила в редакцию 11 февраля 2019

© Белорусский государственный университет информатики и радиоэлектроники, 2019

Аннотация. В работе рассматривается понятие «интеллектуальный анализ данных», описывается концепция облачных вычислений. В качестве примера реализации интеллектуального анализа в виде облачных вычислений рассматривается система, разработанная и развернутая на вычислительном кластере кафедры ЭВМ БГУИР. Приводятся некоторые результаты проведения исследований при помощи данной системы.

Ключевые слова: интеллектуальный анализ данных, облачные вычисления, частное облако, вычислительный кластер, алгоритм k -средних, сегментация изображений, аннотирование изображений.

Конфликт интересов. Авторы заявляют об отсутствии конфликта интересов.

Для цитирования. Татур М.М., Лукашевич М.М., Перцев Д.Ю., Искра Н.А. Интеллектуальный анализ данных и облачные вычисления. Доклады БГУИР. 2019; 6(124): 62-71.

INTELLIGENT DATA ANALYSIS AND CLOUD COMPUTING

TATUR M.M., LUKASHEVICH M.M., PERTSEV D.Y., ISKRA N.A.

Belarusian State University of Informatics and Radioelectronics, Republic of Belarus

Submitted 11 February 2019

© Belarusian State University of Informatics and Radioelectronics, 2019

Abstract. In this paper the term of «intelligent data analysis» is discussed, the cloud computing concept is described. The system developed and deployed on the computer cluster by ECM department of BSUIR is shown as an example of the intelligent data analysis by means of cloud computing. Some results of research with the help of this system are given.

Keywords: intelligent data analysis, cloud computing, private cloud, computer cluster, k -means, image segmentation, image captioning.

Conflict of interests. The authors declare no conflict of interests.

For citation. Tatur M.M., Lukashevich M.M., Pertsev D.Y., Iskra N.A. Intelligent data analysis and cloud computing. Doklady BGUIR. 2019; 6(124): 62-71.

Введение

«Интеллектуальный анализ данных» – это современное научное направление, которое исследует и решает проблемы, связанные с поиском неочевидных закономерностей и причинно-следственных связей в системах, процессах и явлениях. По сути, это получение первичной информации, которая в дальнейшем используется для принятия технических, управленческих и других решений. Близкими к термину «Интеллектуальный анализ данных» являются англоязычные «Data Mining & Knowledge Discovery» и более общее «Machine Learning». Можно считать, что данное направление является составной частью более широкого понятия – «Искусственный интеллект». Онтологию научных направлений можно представить в виде мнемосхемы на рис. 1.

Как видно из рис. 1, область искусственного интеллекта наряду с интеллектуальной обработкой данных включает Логический (в том числе Нечеткий) вывод, Семантическую обработку знаний, оперирующую информацией более высокого уровня, и ряд других направлений.



Рис. 1. Направления исследований в области искусственного интеллекта
Fig. 1. Directions of research in the field of artificial intelligence

Примерами прикладных задач, решаемых в рамках аппарата интеллектуального анализа данных могут служить: распознавание изображений (например, лиц, отпечатков пальцев), автоматическая индексация (разметка) изображений, прогнозирование погоды, котировок курсов валют, анализ причин роста или спада продаж и т. п. В работах [1–3] описаны решения некоторых из них. Все задачи подобного рода сопряжены с анализом большого объема эмпирических данных. Поэтому это направление тесно связывают с понятиями «Big Data» и «Data Analysis». Как показано в работах [4–6], вычислительная сложность алгоритмов интеллектуального анализа возрастает экспоненциально, в зависимости от объема данных. Это приводит к необходимости применения вычислительных средств с параллельной архитектурой.

На кафедре ЭВМ БГУИР активно ведутся научно-исследовательские работы в отмеченных направлениях. Развернута лаборатория высокопроизводительных вычислений с удаленным доступом пользователей. По сути, организовано облако, в котором реализована оригинальная система интеллектуального анализа данных. Система используется в учебно-исследовательских целях и позволяет клиентам с минимальными трудозатратами решать практические задачи интеллектуального анализа данных. В настоящей работе изложены теоретические основы и практические принципы построения рассматриваемой системы, а также приведены примеры решения некоторых прикладных задач, что представляет собой комплексный результат труда коллектива кафедры ЭВМ за последнее пятилетие.

Методология интеллектуального анализа данных

Решение прикладных задач интеллектуального анализа данных в общем случае сводится к последовательности действий, показанных на рис. 2 [7–9]. Математическую основу, или ядро системы, образуют формальные алгоритмы обработки (например, сравнение, сопоставление) упорядоченных наборов данных. В литературе такие данные также называют вектором информативных признаков или образом (англ. pattern). Сами алгоритмы в достаточной степени известны и реализованы в различных библиотеках и фреймворках,

о чем более детально будет сказано ниже. Все остальные блоки (предобработка, постобработка и обучение) в той или иной степени обеспечивают функционирование ядра.

При решении прикладной задачи исследователю предстоит выполнить следующие действия: определить, какие исходные данные использовать, как их нормализовать и взвешивать, какие формальные алгоритмы анализа данных применять и в какой последовательности, а также оценить качество и достоверность полученного результата. В целом задача интеллектуального анализа данных остается неформальной, а роль исследователя состоит в обучении системы, т. е. корректировке всех параметров в зависимости от качества получаемых результатов. В некоторых частных случаях прикладная задача, решенная с участием разработчика в заданных ограничениях, может в дальнейшем тиражироваться с автоматическим обучением. При этом создается ошибочное впечатление о полном исключении исследователя из процесса анализа данных.

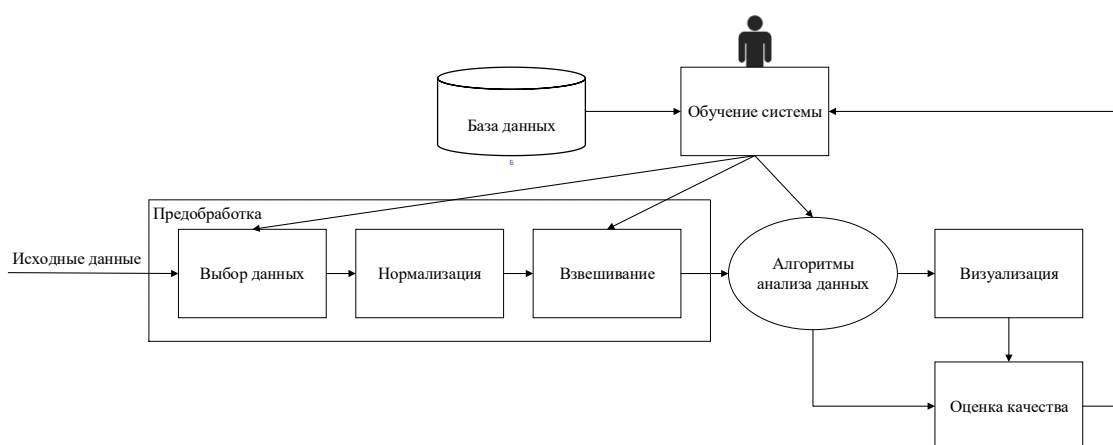


Рис. 2. Обобщенная функциональная схема интеллектуального анализа данных
Fig. 2. Generalized functional diagram of data mining

Таким образом, изначально, решение прикладной задачи в области интеллектуального анализа данных – это процесс проведения исследования с итеративной постановкой экспериментов, связанных с выдвижением и проверкой гипотез. При этом анонсируемая система представляет собой инструментальное средство для проведения исследований, призванное повысить качество создаваемых интеллектуальных систем и сократить сроки их разработки.

Понятие облачных вычислений

Традиционная модель организации вычислений предполагает создание инфраструктуры (установку и настройку серверов, организацию хранилища данных, организацию локальной сети, связывающей всю инфраструктуру в единую сеть), установку и настройку операционной системы, необходимых системных приложений (например, драйверов, платформы .NET, Java, сторонних системных библиотек) и, наконец, установку необходимого прикладного программного обеспечения (рис. 3), с которым будет взаимодействовать конечный пользователь. Очевидно, что, следуя данной модели, необходимо выделить бюджет для приобретения в собственность оборудования, программного обеспечения, оплатить работу специалистов и далее поддерживать все это в актуальном состоянии.

Альтернативой традиционному подходу является модель облачных вычислений (англ. cloud computing), которая состоит в организации удаленного доступа к некоторому общему фонду конфигурируемых вычислительных ресурсов, которые могут быть оперативно предоставлены с минимальными эксплуатационными затратами [10, 11]. Данная модель предполагает существование организации-провайдера, который предоставляет услуги, связанные с выделением необходимой инфраструктуры, поддержанием ее в актуальном состоянии и, возможно, установкой необходимого программного обеспечения.

Облачные вычисления принято классифицировать по двум основаниям [12]: в соответствии с моделью предоставления услуг; в соответствии с моделью развертывания.

В соответствии с моделью предоставления услуг выделяются следующие подходы:

- инфраструктура как сервис (Infrastructure as a Service, IaaS);
- платформа как сервис (Platform as a Service, PaaS);
- программное обеспечение как сервис (Software as a Service, SaaS).



Рис. 3. Традиционная модель организации вычислений
Fig. 3. The traditional model of computing

Различия между указанными моделями заключается в предоставляемом уровне сервисов и услуг (рис. 4).

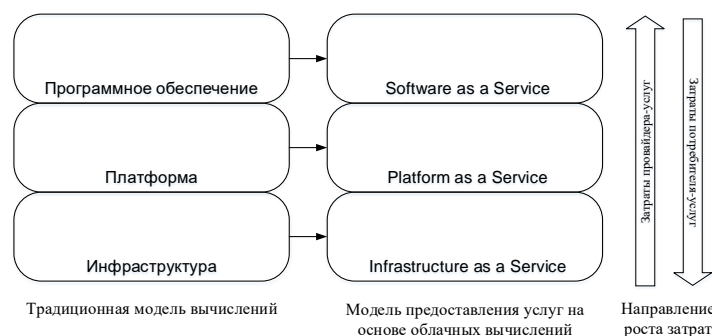


Рис. 4. Соответствие традиционной модели вычислений и услуг, предоставляемых на основе облачных вычислений
Fig. 4. Compliance with the traditional cloud computing model and services

Модель IaaS предполагает предоставление только вычислительных ресурсов в соответствии с запросами потребителя. В дальнейшем пользователь вправе использовать предоставленные ресурсы так, как он считает нужным. Модель PaaS совершенствует подход, предлагаемый в IaaS, и включает уже установленную и подготовленную к работе операционную систему. Модель SaaS является наиболее трудоемкой для провайдера услуг и предполагает создание собственных сервисов и предоставление к ним постоянного одновременного доступа для множества клиентов.

В соответствии с моделью развертывания выделяются следующие подходы: частное облако; публичное облако; гибридное облако.

Ввиду своей открытости, публичное облако предоставляет максимальный спектр возможностей. Однако для организации облака такого уровня требуются существенные вложения в вычислительные ресурсы, которые будут распределяться между множеством клиентов, и решение вопросов, связанных с безопасностью (например, хакерские атаки, изоляция пользовательских данных и т. п.).

Частное облако, как правило, разворачивается организацией для собственных нужд и не предполагает полноценный доступ к системе всем желающим. В случае необходимости организация доступа к облаку, территориально располагаемому за пределами организации, как правило, формируется через закрытую и недоступную извне сеть. Основными преимуществами данного подхода являются пониженные требования к организации защиты доступа, пониженные требования к инфраструктуре.

Гибридное облако является объединением идеологий частного и публичного облаков. С одной стороны, предполагается наличие собственного частного облака. При этом из-за нехватки вычислительных ресурсов или из-за необходимости получения доступа к сторонним ресурсам, организация дополнительно арендует услуги, предоставляемые сторонними публичными облаками. Как правило, в этом случае вся имеющаяся инфраструктура (собственная и арендуемая) объединяется в единую.

Вычислительный кластер БГУИР как провайдер услуг облачных вычислений

В результате модернизации системы в 2019 году вычислительный кластер БГУИР будет являться частным облаком, построенным на основе модели SaaS и включающем 7 Blade-серверов, один из которых выделен как управляющий и 6 в качестве вычислительных, соединенных по шине InfiniBand 4x QDR (40 Gbps). На каждом вычислительном модуле установлено два Intel Xeon E5-2650, 32 Gb RAM стандарта DDR3 и две NVIDIA Tesla M2075 6 Gb RAM.

В качестве услуг, к которым в ближайшей перспективе планируется предоставить доступ, выбраны учебно-исследовательская система интеллектуального анализа данных, разрабатываемая сотрудниками кафедры ЭВМ и НИЛ 3.6 [13–15], а также Apache Zeppelin [16].

Основными преимуществами использования собственного вычислительного кластера БГУИР в качестве вычислительного облака являются:

- независимость от внешних организаций;
- подключение к локальной сети БГУИР, что позволяет использовать кластер студентами и исследователями, в том числе магистрантами и аспирантами, на любой кафедре при получении логина и пароля. В перспективе имеется возможность организации доступа через Интернет, что позволит подключать к системе любого желающего;
- достаточные вычислительные ресурсы для того, чтобы справиться с серьезной вычислительной нагрузкой, в том числе длительной.

Сервисы, предоставляемые на базе вычислительного кластера БГУИР

Учебно-исследовательская система интеллектуального анализа данных. Разрабатываемая система представляет собой комплекс микросервисов, взаимодействующих между собой (рис. 5): уровень сервисов, включающий библиотеки алгоритмов; уровень алгоритмов анализа данных; Web-сайт (интерфейс пользователя).

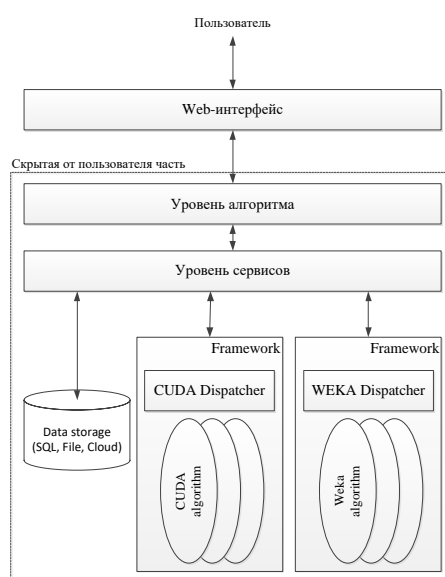


Рис. 5. Структурная схема системы интеллектуального анализа данных

Fig. 5. Block diagram of a data mining system

Уровень сервисов. Существует множество различных библиотек, реализующих алгоритмы интеллектуального анализа данных. Наиболее известные из них: scikit-learn [17], MLlib [18], Theano [19], Weka [20].

Каждая из представленных библиотек (англ. framework) предоставляет собственный интерфейс доступа, поддерживает разные языки программирования. При этом каждый из перечисленных framework'ов эффективно справляется с определенной задачей. В случае если программисту необходимо использовать алгоритмы из разных библиотек, это становится затруднительным и требует высокой квалификации разработчика. Для решения поставленной задачи был разработан первый уровень – уровень сервисов.

Основными задачами уровня сервисов являются:

- подключение запрашиваемой библиотеки либо информирование о неполадках в настройках;
- проверка корректности входных параметров;
- вызов необходимых алгоритмов и обработка исключений;
- предоставление информации об исполнении алгоритмов, в том числе о возникших ошибках.

Уровень алгоритмов анализа данных. Вышележащий уровень – прослойка для организации доступа к сервисам. Его основными задачами являются:

- хранение информации о подключенных библиотеках алгоритмов;
- хранение информации о поддерживаемых алгоритмах в рамках каждой из подключенных библиотек;
- обобщение и систематизация полученной информации;
- взаимодействие с Web-сайтом.

Web-сайт. Конечный уровень – Web-сайт, развернутый на сервере на базе кафедры ЭВМ БГУИР.

Его основными задачами являются:

- удобный доступ к поддерживаемым алгоритмам интеллектуального анализа данных;
- формирование цепочки алгоритмов с необходимыми настройками для решения конкретных задач;
- предоставление полученных результатов работы;
- контроль прав доступа.

Apache Zeppelin. Проект Apache Zeppelin – полностью открытый проект, построенный по принципу Web-блокнотов и созданный для интерактивного анализа данных.

Основным преимуществом данного проекта является полноценная поддержка сложной инфраструктуры, создаваемой для облачных вычислений. На базе вычислительного кластера БГУИР и сервиса Apache Zeppelin предоставляется полноценный доступ через Web-интерфейс к таким языкам программирования, как Python и R. Кроме того, поддерживаются технологии Apache Spark и TensorFlow, созданные для анализа данных и работы с нейронными сетями соответственно.

Примеры решения учебных и прикладных задач

Задача кластеризации алгоритмом k -средних. В качестве простой тестовой задачи для учебно-исследовательской системы анализа данных использовался алгоритм кластеризации k -средних [22]. Аппаратная реализация – вычислительный кластер БГУИР.

Для проведения эксперимента генерировался тестовый набор из 10^7 образов на 100 признаков (координат) с заданным числом кластеров. Из данного набора создавались подмножества с меньшим числом объектов, кратным 10^6 , для исследования зависимости времени от объема обрабатываемых данных.

Ниже представлены полученные зависимости среднего времени обработки от числа узлов кластера (рис. 6, *a*) и от числа объектов (образов) для кластеризации при числе вычислительных узлов – 7 (рис. 6, *b*).

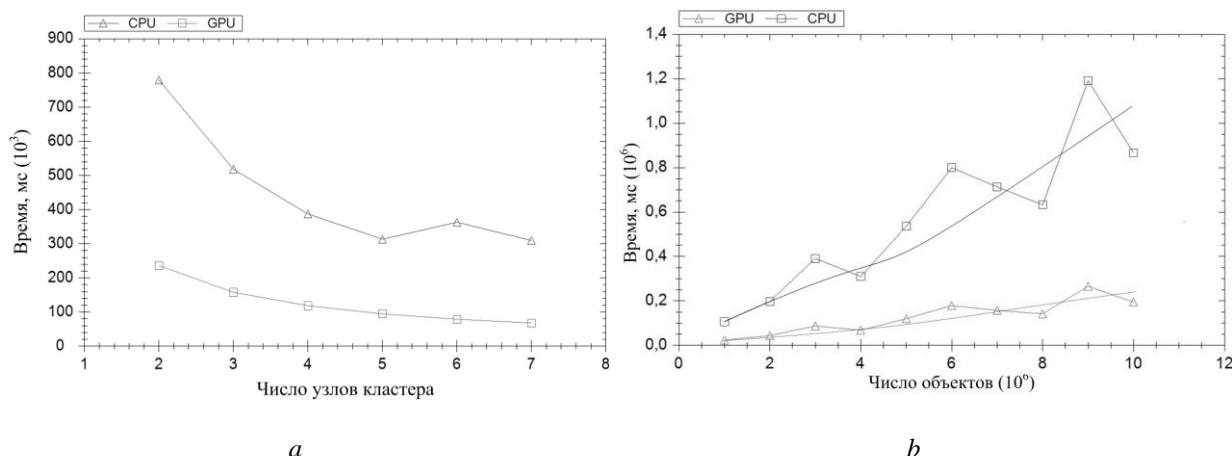


Рис. 6. Зависимости времени обработки от *a* – числа узлов кластера; *b* – числа объектов
Fig. 6. Processing time dependencies *a* – is the number of cluster nodes; *b* – number of objects

Сегментация медицинских изображений. Сегментация изображений – это процесс разделения изображения на составляющие части или объекты. Автоматизированная сегментация медицинских изображений является сложной задачей из-за большого разнообразия форм и размеров объектов предметной области [22].

Сверточная сеть Unet [23] была разработана с учетом особенностей обработки медицинских изображений. Она позволяет достичь высокой точности и использует небольшой набор данных для обучения.

На основе датасета с медицинскими изображениями, который предоставлялся в рамках конкурса «Data Science Bowl 2018», обучалась сеть с Unet-архитектурой. Для анализа работы сети были проведены эксперименты с изменением алгоритмов оптимизации, предоставляемых сервисом Apache Zeppelin. Результаты экспериментов приводятся на рис. 7.

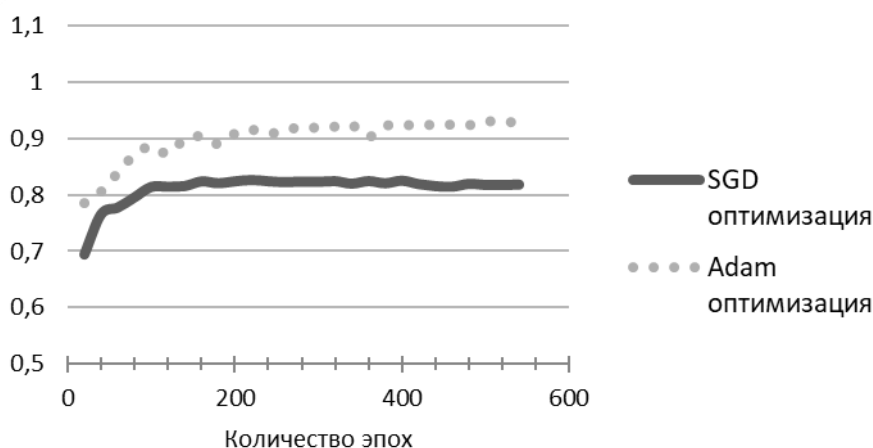


Рис. 7. Изменение коэффициента Жаккара при различных видах оптимизации
Fig. 7. Change in the Jacquard coefficient for various types of optimization

Аннотирование изображений. Решение задачи понимания и интерпретации изображений требует, с одной стороны, точной и быстрой детекции объектов на изображении, а с другой – построения семантического описания изображения на ограниченном естественном языке. Таким образом, ставится задача аннотирования изображений (англ. Image Captioning) – получения словесного описания содержания изображения.

В рамках совместного исследования кафедры ЭВМ и кафедры ИИТ БГУИР [24] был рассмотрен подход к семантическому анализу изображений на основе интеграции модели,

использующей сверточные нейронные сети, и средств представления и обработки информации в рамках открытой семантической технологии проектирования интеллектуальных систем.

При помощи сервиса Apache Zeppelin можно построить цепочки алгоритмов обработки данных. Для решения задачи аннотирования изображений была предложена последовательность обработки с обратными связями, представленная на рис. 8.

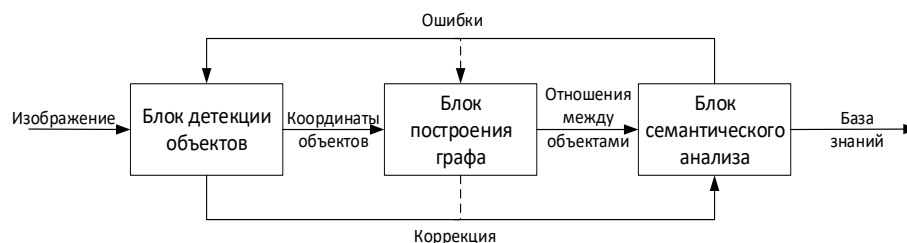


Рис. 8. Архитектура обработки изображения для аннотирования
Fig. 8. Image processing architecture for annotation

Блок детекции построен на основе архитектуры Faster R-CNN [25], реализация внутренних алгоритмов которой предоставляется через Apache Zeppelin, взаимодействующей с алгоритмами TensorFlow. Результаты детекции передаются в блок построения графа, который учитывает координаты расположения выделенных объектов. Далее отношения между связанными объектами уточняются. Блок семантического анализа, построенный в виде OSTIS-системы, на основании характера выделенных объектов определяет контекст изображения. Данный блок определяет возможные ошибки либо в детекции объектов, либо в отношениях. Блок детекции (или блок построения графа, в зависимости от характера ошибки) перепроверяет результаты для «подозрительных» регионов, модель корректируется. По скорректированной модели, принимая во внимания языковые отношения «субъект-объект», обозначения отношений заменяются подходящими языковыми конструкциями. Результаты в дальнейшем могут интегрироваться в базу знаний и использоваться для последующего улучшения семантического анализа.

Заключение

В представленной работе обобщены основные результаты научно-исследовательской деятельности в контексте интеллектуального анализа данных и облачных вычислений. Представлена обновленная версия вычислительного кластера БГУИР с открывающимися новыми возможностями по предоставлению услуг по анализу и обработке больших объемов данных, в том числе по обработке изображений.

Учебно-исследовательская система интеллектуального анализа данных создавалась при финансовой поддержке в рамках государственной программы научных исследований «Информатика, космос и безопасность», подпрограмма 1 «Информатика и космические исследования» (задание 1.6.04 «Разработка интеллектуальной вычислительной системы обработки больших объемов данных»).

Список литературы

1. Поиск, визуализация скрытых зависимостей и прогнозирования развития ситуаций на базе технологий Data Mining & Knowledge Discovery / М.М. Татур [и др.] // Сб. матер. конф. «BIG DATA and Advanced Analytics». Минск, 2016. С. 194–196.
2. Живицкая Е.Н., Пархименко В.А., Татур М.М. Технологии Data Mining & Knowledge Discovery в принятии решений в области маркетинга, менеджмента и логистики // Сб. матер. конф. «Сучасні проблеми і досягнення в галузі радіотехніки, теле-комунікацій та інформаційних технологій». Запоріжжє, 2016. С. 326–329.
3. Применение методов DataMining и Knowledge Discovery в оперативно-розыскной деятельности / С.Н. Нефедов [и др.] // Сб. матер. конф. «Актуальные проблемы оперативно-розыскной деятельности». Минск, 2017. С. 70–72.
4. Татур М.М. Особенности построения вычислителей интеллектуальной обработки данных //

- Информатика. 2015. № 1 (45). С. 39–44.
5. К вопросу о подготовке данных для решения задач Data Mining / Е. Н. Живицкая [и др.] // Сб. матер. конф. «BIG DATA and Advanced Analytics». Минск, 2017. С. 288–292.
 6. Интеллектуальный анализ данных: trend или application? / М.М. Татур [и др.] // Сб. матер. конф. «Информационные технологии и системы». Минск, 2017. С. 10–12.
 7. Tatur M.M., Iskra N.A. Intelligent Data Analysis: From Theory to Practice // Сб. матер. конф. «Открытые семантические технологии проектирования интеллектуальных систем». Минск, 2018. С. 171–175.
 8. Мюллер А., Гвидо С. Введение в машинное обучение с помощью Python: руководство для специалистов по работе с данными. М., 2017. 393 с.
 9. Силен Д., Мейсман А., Али М. Основы Data Science и Big Data. Python и наука о данных. СПб.: Питер, 2017. 336 с.
 10. Mell P., Grance T. The NIST Definition of Cloud Computing / Recommendations of the National Institute of Standards and Technology. NIST, 2011.
 11. Marinescu D.C. Cloud Computing: Theory and Practice. Morgan Kaufmann, 2017. 588 p.
 12. Bhowmik S. Cloud Computing. Cambridge University Press, 2017. 426 p.
 13. Демидчук А.И., Перцев Д.Ю., Самаль Д.И. Учебно-исследовательская система обработки больших данных // BIG DATA and Advanced Analytics. Минск: БГУИР, 2017. С. 170–173.
 14. Система обработки больших данных на основе вычислительного кластера БГУИР / Д. И. Самаль [и др.] // Сб. матер. конф. «BIG DATA Advanced Analytics». Минск, 2018. С. 220–256.
 15. Интеллектуальная обработка больших объемов данных на основе технологий MPI и CUDA. Лабораторный практикум : пособие / А. И. Демидчук [и др.]. Минск : БГУИР, 2017. 60 с.
 16. Zeppelin [Электронный ресурс]. URL: <http://zeppelin.apache.org/> (дата обращения: 20.01.2019).
 17. Scikit-learn: Machine Learning in Python [Электронный ресурс]. URL: <https://scikit-learn.org/stable/> (дата обращения: 20.01.2019).
 18. MLlib Apache Spark [Электронный ресурс]. URL: <https://spark.apache.org/mllib/> (дата обращения: 20.01.2019).
 19. Theano 1.0.0 documentation [Электронный ресурс]. URL: <http://deeplearning.net/software/theano/> (дата обращения: 20.01.2019).
 20. Weka 3 – Data Mining with Open Source Machine Learning Software in Java [Электронный ресурс]. URL: <https://www.cs.waikato.ac.nz/ml/weka/> (дата обращения: 20.01.2019).
 21. Проявление закона Амдала–Густавсона на примере реализации алгоритма k-средних / А. И. Демидчук [и др.] // Сб. матер. конф. «BIG DATA and Predictive Analytics». Минск, 2015. С. 151–154.
 22. Лукашевич М.М., Старовойтов В.В. Методика подсчета числа ядер клеток на медицинских гистологических изображениях // Системный анализ и прикладная информатика. 2016. № 2. С. 38–42.
 23. Ronneberger O., Fischer P., Brox T. U-Net: Convolutional Networks for Biomedical Image Segmentation / Medical Image Computing and Computer-Assisted Intervention, 2015. P. 234–241.
 24. Семантическая модель представления и обработки баз знаний / В.В. Голенков [и др.]. // Аналитика и управление данными в областях с интенсивным использованием данных. 2017. С. 412–419.
 25. Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks / Shaoqing R. [et al.] // IEEE Transactions on Pattern Analysis and Machine Intelligence. 2017. P. 1137–1149.

References

1. Poisk, vizualizacija skrytyh zavisimostej i prognozirovanija razvitija situacij na baze tehnologij Data Mining & Knowledge Discovery / М.М. Татур [и др.] // Sb. mater. konf. «BIG DATA and Advanced Analytics». Minsk, 2016. S. 194–196. (in Russ.)
2. Zhivickaja E.N., Parhimenko V.A., Tatur M.M. Tehnologii Data Mining & Knowledge Discovery v prinjatii reshenij v oblasti marketinga, menedzhmenta i logistiki // Sb. mater. konf. «Suchasni problemi i dosjagnennja v galuzi radiotekhniki, tele-komunikacij ta informacijnih tehnologij». Zaporozh'e, 2016. S. 326–329. (in Russ.)
3. Primenenie metodov DataMining i Knowledge Discovery v operativno-rozysknoj dejatel'nosti / S.N. Nefedov [I dr.] // Sb. mater. konf. «Aktual'nye problemy operativno-rozysknoj dejatel'nosti». Minsk, 2017. S. 70–72. (in Russ.)
4. Tatur M.M. Osobennosti postroenija vychislitelej intellektual'noj obrabotki dannyh // Informatika. 2015. № 1 (45). S. 39–44. (in Russ.)
5. K voprosu o podgotovke dannyh dlja reshenija zadach Data Mining / E. N. Zhivickaja [i dr.] // Sb. mater. konf. «BIG DATA and Advanced Analytics». Minsk, 2017. S. 288–292. (in Russ.)
6. Intellektual'nyj analiz dannyh: trend ili application? / M.M. Tatur [i dr.] // Sb. mater. konf. «Informacionnye tehnologii i sistemy». Minsk, 2017. S. 10–12. (in Russ.)
7. Tatur M.M., Iskra N.A. Intelligent Data Analysis: From Theory to Practice // Sb. mater. konf. «Otkrytye semanticheskie tehnologii proektirovanija intellektual'nyh sistem». Minsk, 2018. S. 171–175.

8. Mjuller A., Gvido S. Vvedenie v mashinnoe obuchenie s pomoshh'ju Python: Rukovodstvo dlja specialistov po rabote s dannymi. M., 2017. 393 s. (in Russ.)
9. Silen D., Mejsman A., Ali M. Osnovy Data Science i Big Data. Python i nauka o dannyh. SPb.: Piter, 2017. 336 s. (in Russ.)
10. Mell P., Grance T. The NIST Definition of Cloud Computing / Recommendations of the National Institute of Standards and Technology. NIST, 2011.
11. Marinescu D.C. Cloud Computing: Theory and Practice. Morgan Kaufmann, 2017. 588 p.
12. Bhowmik S. Cloud Computing. Cambridge University Press, 2017. 426 p.
13. Demidchuk A.I., Percev D.Ju., Samal' D.I. Uchebno-issledovatel'skaja sistema obrabotki bol'shikh dannyh // BIG DATA and Advanced Analytics. Minsk: BGUIR, 2017. S. 170–173. (in Russ.)
14. Sistema obrabotki bol'shikh dannyh na osnove vychislitel'nogo klastera BGUIR / D.I. Samal' [i dr.] // Sb. mater. konf. «BIG DATA Advanced Analytics». Minsk, 2018. S. 220–256. (in Russ.)
15. Intellektual'naja obrabotka bol'shikh ob'emov dannyh na osnove tehnologij MPI i CUDA. Laboratornyj praktikum: posobie / A. I. Demidchuk [i dr.]. Minsk: BGUIR, 2017. 60 s. (in Russ.)
16. Zeppelin [Electronic resource]. URL: <http://zeppelin.apache.org/> (data obrashhenija: 20.01.2019).
17. Scikit-learn: Machine Learning in Python [Electronic resource]. URL: <https://scikit-learn.org/stable/> (date of access: 20.01.2019).
18. MLib Apache Spark [Electronic resource]. URL: <https://spark.apache.org/mllib/> (date of access: 20.01.2019).
19. Theano 1.0.0 documentation [Electronic resource]. URL: <http://deeplearning.net/software/theano/> (date of access: 20.01.2019).
20. Weka 3 – Data Mining with Open Source Machine Learning Software in Java [Electronic resource]. URL: <https://www.cs.waikato.ac.nz/ml/weka/> (date of access: 20.01.2019).
21. Projavlenie zakona Amdala-Gustavsona na primere realizacii algoritma k-srednih / A.I. Demidchuk [i dr.] // Sb. mater. konf. «BIG DATA and Predictive Analytics». Minsk, 2015. S. 151–154. (in Russ.)
22. Lukashevich M.M., Starovojtov V.V. Metodika podscheta chisla jader kletok na medicinskih gistologicheskikh izobrazhenijah // Sistemnyj analiz i prikladnaja informatika. 2016. № 2. S. 38–42. (in Russ.)
23. Ronneberger O., Fischer P., Brox T. U-Net: Convolutional Networks for Biomedical Image Segmentation / Medical Image Computing and Computer-Assisted Intervention, 2015. P. 234–241. (in Russ.)
24. Semanticheskaja model' predstavlenija i obrabotki baz znaniy / V.V. Golenkov [i dr.]. // Analitika i upravlenie dannymi v oblastjakh s intensivnym ispol'zovaniem dannyh. 2017. S. 412–419. (in Russ.)
25. Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks / Shaoqing R. [et al.] // IEEE Transactions on Pattern Analysis and Machine Intelligence. 2017. P. 1137–1149.

Сведения об авторах

Татур М.М., д.т.н., профессор, профессор кафедры электронных вычислительных машин Белорусского государственного университета информатики и радиоэлектроники.

Лукashevich M.M., к.т.н., доцент, декан факультета компьютерных систем и сетей Белорусского государственного университета информатики и радиоэлектроники.

Перцев Д.Ю., научный сотрудник НИЛ 3.6 НИЧ Белорусского государственного университета информатики и радиоэлектроники.

Искра Н.А., старший преподаватель кафедры электронных вычислительных машин Белорусского государственного университета информатики и радиоэлектроники.

Адрес для корреспонденции

220013, Республика Беларусь,
г. Минск, ул. П. Бровки, 6,
Белорусский государственный университет
информатики и радиоэлектроники
тел. +375-29-586-93-52;
e-mail: niskra@bsuir.by
Искра Наталья Александровна

Information about the authors

Tatur M.M., D.Sci, professor of the department of electronic computing machines of Belarusian State University of Informatics and Radioelectronics.

Lukashevich M.M., PhD, associate professor, dean of the faculty of computer systems and networks of Belarusian State University of Informatics and Radioelectronics.

Pertsev D.Y., researcher of SRL 3.6 of R&D department of Belarusian State University of Informatics and Radioelectronics.

Iskra N.A., senior lecturer of the department of electronic computing machines of Belarusian State University of Informatics and Radioelectronics.

Address for correspondence

220013, Republic of Belarus,
Minsk, P. Brovka, st., 6,
Belarusian State University
of Informatics and Radioelectronics
tel. +375-29-586-93-52;
e-mail: niskra@bsuir.by
Iskra Natalia Alexandrovna