

# АВТОМАТИЗИРОВАННЫЙ АНАЛИЗ ДАННЫХ В МЕДИЦИНСКИХ ПРИЛОЖЕНИЯХ ИНТЕРНЕТА ВЕЩЕЙ

В. В. Ефимов<sup>1\*</sup>, Р. Я. Пирмагомедов<sup>1</sup>

<sup>1</sup> СПбГУТ, Санкт-Петербург, 193232, Российская Федерация  
\* Адрес для переписки: quterman@gmail.com

## **Аннотация**

Технологические решения на основе концепции Интернета Вещей уже давно вышли за границы развлекательной сферы и находят применение в таких важных отраслях как управление промышленным производством, транспортом и здравоохранение. В здравоохранении, использование устройств Интернета Вещей, предоставляет пациентам и врачам доступ к данным о состоянии здоровья в режиме реального времени и в удобном формате. В совокупности, применение Интернета Вещей в медицине позволяет быстро и качественно реагировать на любые изменения в функциональном состоянии организма и симптомы пациента. Вследствие всепроникающей информатизации и внедрения новых средств диагностики, возникают большие массивы данных, требующие анализа. В последние годы были проведены множественные исследования, нацеленные на сбор данных их систематизацию и анализ. Однако, все отрасли имеют свою уникальную специфику и требуют индивидуального подхода при внедрении.

В этой статье рассматриваются медицинские приложения Интернета Вещей и проблемы анализа данных, которые требуют решения при внедрении подобных приложений на практике. В частности, рассматриваются вопросы автоматизированного анализа больших данных с использованием облачных вычислений на основе Hadoop.

## **Ключевые слова**

Интернет Вещей, большие данные, анализ данных.

## **Информация о статье**

УДК 004.75

Язык статьи – русский.

Поступила в редакцию 28.06.18, принята к печати 03.09.18.

**Ссылка для цитирования:** Ефимов В. В., Пирмагомедов Р. Я. Автоматизированный анализ данных в медицинских приложениях Интернета Вещей // Информационные технологии и телекоммуникации. 2018. Том 6. № 3. С. 73–78.

# DEVELOPMENT OF ALGORITHMS FOR AUTOMATED DATA ANALYSIS IN MEDICAL APPLICATIONS INTERNET OF THINGS

V. Efimov<sup>1\*</sup>, R. Pirmagomedov<sup>1</sup>

<sup>1</sup> SPbSUT, St. Petersburg, 193232, Russian Federation

\* Corresponding author: quterman@gmail.com

**Abstract**—Technological solutions based on the Internet of Things concept has gone beyond the boundaries of the entertainment sector and are used in such mission-critical sectors as industrial production management, transport, and healthcare. In healthcare, IoT devices, provides access to health data in real time and in a convenient format. In total, the use of the Internet of things in healthcare allows one to quickly and efficiently respond any changes in the functional state of the body and the patient's symptoms. Due to the pervasive Informatization and deployment of new diagnostic tools, there are large amounts of data requires analysis. In recent years, numerous studies on the collection and analysis of the Big Data. However, all industries have their own unique specifics and require individual implementation approaches.

This article discusses the medical applications of the Internet of Things and the problems of data analysis that need to be addressed while implementation of such applications in practice. In particular, we consider the issues of automated analysis of big data using cloud computing based on Hadoop.

**Keywords**—Internet of things, big data, data science.

## Article info

Article in Russian.

Received 28.06.18, accepted 03.09.18.

**For citation:** Efimov V., Pirmagomedov R.: Development of algorithms for automated data analysis in medical applications Internet of Things // Telecom IT. 2018. Vol. 6. Iss. 3. pp. 73–78 (in Russian).

## Введение

Интернет вещей (*Internet of Things*, IoT) – это глобальная инфраструктура для информационного общества, которая позволяет предоставлять более сложные услуги путем соединения между собой вещей на основе существующих и развивающихся информационных технологий. Новые технологические возможности, которые стали возможными благодаря IoT, находят применение во множестве сфер человеческой деятельности от развлечений до медицины и промышленности. Внедрение IoT технологий позволяет автоматизировать процессы повысив тем самым их эффективность.

В этой статье речь пойдет о медицинских приложениях Интернета Вещей и проблемах, которые они помогут решить. Использование датчиков, медицинских приборов, работающих удаленно, предоставляет пациентам и врачам доступ к данным о состоянии здоровья в режиме реального времени и в удобном формате [1]. В совокупности, применение интернета вещей в медицине позволяет

быстро и качественно реагировать на любые изменения в функциональном состоянии организма и симптомы пациента.

### **Проблемы развития медицинских приложений Интернета Вещей**

Несмотря очевидные преимущества от внедрения IoT технологий в медицине, на пути к автоматизации возникает больше количества препятствий, в частности – это проблема большого объема данных, который генерируют датчики. Возникает необходимость постоянного набора персонала для обработки и контроля этих показаний [2]. Эта проблема требует появления комплексного решения, которое сможет выполнять непрерывный сбор данных, контроль параметров среды, реакций пациентов и других разнообразных процессов. Но самым главным требованием становится обработка «сырых» медицинских данных, поступающих с различных источников, которые нуждаются в быстрой обработке и преобразовании в удобные для специалиста показатели, ориентируясь на которые можно будет с легкостью следить за состоянием пациента без личного контакта с ним и без глубокого погружения в собранную системой информацию.

Медицинская область включает в себя огромное количество данных, которое требует качественного анализа и структурирования. За последние годы было проведено множество исследований в данной области, нацеленных на сбор данных, их анализ и визуализацию.

При непрерывной информатизации сферы здравоохранения, объем данных увеличивается. По общей статистики из EMC было определено, что в 2014 г. объем данных, полученных от медицинских вещей, составлял 4,3 Збайт. И уже к 2020 г. по прогнозам это количество увеличится до 44 Збайт. Говоря про скорость, следует отметить, что данные мониторинга здоровья генерируются с невероятно большой частотой, за каждую секунду может быть произведено несколько измерений, которые необходимо передать [3]. Как правило, медицинские данные включают множество различных источников, например, цифровую медицинскую запись, МРТ, КТ, данные генома и др. Зачастую медицинские данные могут быть неполными, предвзятыми или даже заполненными шумом. Для эффективной работы с данными, требуется предварительное проведение их качественного и количественного анализа автоматизированными средствами.

### **Существующие решения и наработки**

С увеличением объема информации, собранной с помощью систем мониторинга пациентов, врачи стали испытывать все большее давление, чтобы принимать сложные аналитические решения, которые требуют различные типы данных, собираемых на одного пациента. Медицинская обработка данных представляет собой процесс ETL (*Extract, Transform, Load*), интегрированный с некоторыми источниками. В дополнение к методам ETL, широкое распространение получили методы визуализации данных, включая алгоритмы treemap, circle-packing и sunburst, которые являются эффективным инструментом для восприятия специалистом сложных и объемных медицинских данных.

Экспоненциальный рост данных в сфере здравоохранения привел к проблемам их передачи, хранения и обработки. Проблема передачи больших потоков

данных может быть решена с внедрением новых технологий беспроводного доступа 5G NR описанных в спецификациях 3GPP [4]. Для решения проблем хранения и обработки применяются комплексные решения, одним из которых является Hadoop MapReduce. Hadoop – это технология кластеризации вычислительных машин, принцип работы, которой заключается в наличие одной главной машины, на вход которой поступает большое количество данных или вычислительных задач, она разбивает задачу на более простые подзадачи и распределяет её по известным ей кластерам. Кластеры решают подзадачи и возвращают главной машине ответ. Важной особенностью является возможность распределено производить операции приема и обработки данных, процедуры которых могут производиться параллельно.

В работе [5] предложено усовершенствование алгоритмов обработки больших данных на основе MR-MSPolygrap. Главная идея заключалась в распараллеливании потоков данных при помощи MapReduce. Предложенное решение позволяет параллельно производить процессы обработки данных и их свертки. В эксперименте было произведено сравнение работы Polygraph в гибридном (одиночном) режиме и при использовании Hadoop, на основе базы из 64 000 спектров данных. В процессе эксперимента было произведено измерение времени, требуемого на обработку. В результате, данный метод работы с параллельными потоками к значительному увеличению скорости обработки данных. Для обработки спектра Polygraph в гибридном режиме потребовалось 2 000 часов, использование же большого количества процессоров свело это время к 6 часам. По данным эксперимента, предложенный авторами подход позволяет существенно увеличить скорость обработки данных.

Markonis D. [6] в своем исследовании разбирал проблему роста объема медицинских изображений, количество которых также увеличивается с каждым годом. В процессе работы был создан кластер узлов с использованием Hadoop, позволяющий выполнять до 42 одновременных задач и рассмотрена проблема индексации медицинских изображений. В результате было предложено 2 подхода к индексации: компонентная и монолитная. Компонентная стратегия была удобна тем, что для обработки не требовалось запуск всей сети машин, однако MapReduce плохо справлялся с задачами ввода-вывода [7]. Монолитная стратегия показала себя значительно лучше, при исследовании с помощью нее удалось произвести индексирования 100 000 изображений примерно за час.

## Заключение

На сегодняшний день перед архитекторами и разработчиками приложений, реализующих возможности интернета вещей в медицине остро стоит проблема обработки больших данных. Эта проблема обусловлена возникновением все большего количества данных, генерируемых приборами медицинской диагностики. В частности, это врачебные записи, лабораторные отчеты, показания диагностических приборов. Все эти данные бесполезны до тех пор, пока не будут надлежащим способом обработаны. Процесс анализа таких данных, который позволит быстро определять проблемы и помогать врачам оперативно на них реагировать и применять меры, становится важным направлением в развитии алгоритмов интеллектуального анализа данных. В данной статье были рассмотрены примеры работ в данном направлении. Представлены решения с использованием

облачных вычислений на основе Hadoop, позволяющие увеличить скорость обработки данных.

### Литература

1. Глушаков Р. И., Киричек Р. В., Пирмагомедов Р. Я. Медицинские приложения интернета нано-вещей // Актуальные проблемы инфотелекоммуникаций в науке и образовании. VI Международная научно-техническая и научно-методическая конференция: сб. науч. ст. в 4-х т. СПб.: СПбГУТ, 2017. Т. 1. С. 8–14.
2. Блинников М. А., Пирмагомедов Р. Я. Оптимизация нагрузки на сети связи общего пользования, вызванной трафиком медицинских наносетевых приложений Вещей // Информационные технологии и телекоммуникации. 2016. Том 4. № 3. С. 22–30.
3. Pirmagomedov R., Blinnikov M., Glushakov R., Muthanna A., Kirichek R., Koucheryavy A. Dynamic Data Packaging Protocol for Real-Time Medical Applications of Nanonetworks // Internet of Things, Smart Spaces, and Next Generation Networks and Systems, 2017. Pp. 196–205.
4. Борисова М. В., Парамонов А. И., Пирмагомедов Р. Я. Анализ свойств трафика машина-машина и его влияния на качество обслуживания // Распределенные компьютерные и телекоммуникационные сети: управление, вычисление, связь (DCCN-2016): материалы Девятнадцатой международной научной конференции, 21–25 нояб. 2016 г.: в 3 т.; под общ. ред. В. М. Вишневского и К. Е. Самуйлова. М.: РУДН, 2016. С. 36–43.
5. Kalyanaraman A., Cannon W. R., Latt B., and Baxter D. J. Mapreduce implementation of a hybrid spectral library-database search method for large-scale peptide identification // Bioinformatics. 2011. Vol. 27. No. 21. pp. 3072–3073.
6. Markonis D., Schaer R., Eggel I., Muller H., and Depeursinge A., Using "mapreduce for large-scale medical image analysis // IEEE Second International Conference on Healthcare Informatics, Imaging and Systems Biology. HISB 2012, La Jolla, CA, USA, 27–28 September 2012, p. 1.
7. White B., Yeh T., Lin J., and Davis L. Web-scale computer vision using MapReduce for multimedia data mining // Proceedings of the Tenth International Workshop on Multimedia Data Mining, MDMKDD 10, New York, NY, USA, 2010 ACM. pp. 9:1–9:10.

### References

1. Glushakov, R. I., Kirichek, R. V., Pirmagomedov, R. Ya. Medical applications of the Internet Nano-Things // Actual problems of info-telecommunications in science and education. VI international scientific-technical and scientific-methodical conference: sat. art. in 4. St. Petersburg: SPbSUT, 2017. Vol. 1. pp. 8–14.
2. Blinnikov, M. A., Pirmagomedov, R. Ya. Optimization of the load on the communication network caused by traffic of medical applications of nanonetworks // Information technology and telecommunications. 2016. Vol. 4. Iss. 3. pp. 22–30.
3. Pirmagomedov, R., Blinnikov, M., Glushakov, R., Muthanna, A., Kirichek, R., Koucheryavy, A. Dynamic Data Packaging Protocol for Real-Time Medical Applications of Nanonetworks // Internet of Things, Smart Spaces, and Next Generation Networks and Systems, 2017. pp. 196–205.
4. Borisov, M. V., Paramonov, A. I., Pirmagomedov, R. Ya. Analysis of the properties of the traffic is machine-to-machine and its impact on the quality of the service // Distributed computer and communication networks: control, computation, communications (DCCN-2016): proceedings of the Nineteenth international scientific conference, 21–25 Nov. 2016: in 3 volumes; ed. by V. M. Vishnevsky and K. E. Samoylov, M.: PFUR, 2016. pp. 36–43.
5. Kalyanaraman, A., Cannon, W. R., Latt, B., and Baxter, D. J. Mapreduce implementation of a hybrid spectral library-database search method for large-scale peptide identification // Bioinformatics. 2011. Vol. 27. No. 21. pp. 3072–3073.
6. Markonis, D., Schaer, R., Eggel, I., Muller, H., and Depeursinge, A., Using "mapreduce for large-scale medical image analysis // IEEE Second International Conference on Healthcare Informatics, Imaging and Systems Biology. HISB 2012, La Jolla, CA, USA, 27–28 September 2012, p. 1.
7. White, B., Yeh, T., Lin, J., and Davis, L. Web-scale computer vision using MapReduce for multimedia data mining // Proceedings of the Tenth International Workshop on Multimedia Data Mining, MDMKDD 10, New York, NY, USA, 2010 ACM. pp. 9:1–9:10.

**Ефимов Виталий Вячеславович**

– студент магистратуры, СПбГУТ,  
Санкт-Петербург, 193232,  
Российская Федерация, quterman@gmail.com

**Пирмагомедов Рустам Ярахмедович**

– кандидат технических наук, доцент, СПбГУТ,  
Санкт-Петербург, 193232,  
Российская Федерация, lts.pto@yandex.ru

***Efimov Vitaliy***

– Undergraduate, SPbSUT, St. Petersburg,  
193232, Russian Federation,  
quterman@gmail.com

***Pirmagomedov Rustam***

– Candidate of Engineering Sciences,  
Associate Professor, SPbSUT, St. Petersburg,  
193232, Russian Federation, lts.pto@yandex.ru