

Применение облачных технологий в задачах математического анализа кардиологической информации

Е.Ю. Зимина^{1,2}, М.А. Новопашин², А.В. Шмид^{1,2}

¹Национальный исследовательский университет Высшая школа экономики, Московский институт электроники и математики им. А.Н. Тихонова, Таллинская 34, Москва, Россия, 123458

²ЗАО «ЕС-Лизинг», Варшавское шоссе 125, стр. 1, Москва, Россия, 117405

Аннотация. В статье рассматриваются вопросы использования облачных сервисов и технологий в медицинской сфере. Проводится обзор задач математического анализа кардиологической информации с использованием облачных технологий с помощью которых производится хранение, анализ полученных данных и составление дальнейших прогнозов. Рассматривается возможность интеграции облачных технологий с внешними системами.

1. Введение

По прогнозу института McKinsey на 2025 год в сфере цифровой экономики наиболее значимыми областями применения технологий будут Mobile Internet, Automation of knowledge work, Internet of Things и Cloud, суммарный объем рынка которых составит около 30 триллионов долларов. Доля медицинского использования этих технологий оценивается ориентировочно в 30%. Для сравнения, доля нефтегазового сектора мировой экономики оценивается только в 1.5 триллиона долларов.

Также, по сведениям академии Google (Google Scholar) наблюдается резкий рост количества публикаций со ссылкой на Большие данные среди всех публикаций по исследованиям в области здравоохранения.

Целью проекта является создание централизованной системы обработки кардиологической информации населения РФ вне пределов лечебно-профилактических учреждений (ЛПУ), обеспечивающей персонализированный мониторинг, анализ и прогнозирование развития состояния пациента в реальном времени до его первого появления в ЛПУ, а также сопровождение пациентов после выписки из ЛПУ.

Данный проект является одним из возможных ответов на глобальный вызов в области здравоохранения: нарастающим глобальным противоречием между существующими медицинскими возможностями мировой системы здравоохранения с одной стороны и недостатками в концепции и технологиях ее организации, приводящими к быстрому нарастанию нехватки всех видов ресурсов при ее применении, с другой стороны.

Системообразующей компонентой системы обработки кардиологической информации в данном проекте является облако, объединяющее в единую автоматизированную систему

оконечные мобильные устройства, группировку обучаемых компьютеров и медицинский персонал, принимающий решения по сценариям взаимодействия с пациентами.

Разрабатывается система обработки кардиологической информации, обеспечивающая персонафицированный мониторинг, анализ и прогнозирование развития состояния пациента в реальном времени.

2. Облачные технологии

Облако позволяет принимать, хранить и обрабатывать кардиограммы (любого числа отведений), фотоплетизмограммы (пульсовые волны), эхокардиограммы, сейсмокардиограммы для дальнейшего проведения расчётов. Для глубокого анализа требуется наличие первого отведения, но любые дополнительные отведения только уточняют дальнейший анализ.

Облако пополняется посредством application programming interface (API) и адаптеров двух видов: предназначенных для самостоятельной выгрузки данных и предназначенных для чтения форматов ЭКГ, полученных со специализированных устройств.

Облако предназначено для хранения и обработки с помощью технологий Big Data информации, полученной из специализированных учреждений или мобильных устройств пользователя.

Результаты анализа также предоставляются посредством API и специальных адаптеров для интеграции с внешними системами.

Расчеты показателей производятся на основе частотных и энергетических характеристик полученных сигналов. Происходит выделение типовых и переходных частотных состояний для каждого пациента в отдельности, а также для целевых выборок сигналов, разделенных по возрасту, полу и наличию патологий у пациента.

В ходе анализа определяется близость расчетных характеристик пациента к одному из определенных частотных состояний по данным одного обследования. Отслеживается движение показателей состояния пациента между типовыми состояниями самого пациента и целевыми выборками для определения тенденции развития патологий, действия медикаментов, послеоперационной реабилитации.

Определяется биологический возраст сердечно-сосудистой системы пациента, как отдельный вид траектории. Появляется возможность обнаруживать острые состояния пациента, требующих профессиональной помощи.

По удлинению траектории движения расчетных характеристик состояния пациента между типовым состоянием самого пациента и целевыми выборками строится прогноз будущих состояний пациента на ближайшие несколько дней.

При поступлении новой кардиограммы в базу данных (БД) инициируются 2 вида обработки: обработка кардиограммы для текущего пациента и обработка обновленного множества всех кардиограмм из БД.

3. Обработка кардиограмм

Обработка новой поступившей кардиограммы производится в последовательном выполнении нескольких видов заданий:

- подготовка кардиограммы к дальнейшим расчетам;
- расчёт спектра Фурье и его характеристик (гармоники и энергии) для нефильтрованного сигнала, сигнала с исправленной изолинией и сигнала с исправленной изолинией и очищенного от дефектов;
- расчет формы спектра для сигнала с исправленной изолинией и очищенного от дефектов;
- перерасчет кластеров кардиограмм во всей серии кардиограмм пациента с различными параметрами кластеризации;
- расчет показателей тенденции пациента (биологического возраста).

Обработка обновленного множества всех кардиограмм производится для различных целевых групп, сформированных по возрасту, полу и патологиям (кластеризация). Обновление множества кардиограмм выполняется по заданному расписанию.

По патологиям выделяются следующие группы:

- размеченная выборка кардиограмм по диагнозу ишемия (подтвержденная и не подтвержденная);
- размеченная выборка кардиограмм по диагнозу туберкулез (подтвержденная и не подтвержденная);
- размеченная выборка кардиограмм по диагнозу диабет (подтвержденная и не подтвержденная).

Если в группе с момента выполнения последней кластеризации не появилось новых кардиограмм, новая кластеризация не производится.

4. Подготовка кардиограммы для расчетов показателей

Перед обработкой отдельной кардиограммы необходимо удалить дрейф изолинии, и очистить её от сильных дефектов.

Дрейф изолинии удаляется двухмедианным методом, для этого из исходного сигнала поэлементно вычитается рассчитанная посредством двух медиан изолиния.

4.1. Фильтр дефектов

Из сигнала вырезаются RR-интервалы, в которых разброс амплитуд сигнала значительно превышает значение медианы разброса амплитуд по некоторому скользящему окну. Части сигнала склеиваются по вершинам R-пиков, между которыми был вырезан RR-интервал. Краевые части до первого обнаруженного R-пика и после последнего обнаруженного R-пика также отбрасываются, чтобы в итоге все отфильтрованные сигналы начинались с R-пика и заканчивались R-пиком.

Доля отброшенной части сигнала в процентах по отношению ко всей длине сигнала сохраняется отдельно.

5. Расчет спектра Фурье и его характеристик

Для нефильтрованного сигнала, сигнала с исправленной изолинией и сигнала с исправленной изолинией и отфильтрованными дефектами применяется алгоритм быстрого преобразования Фурье. Из комплексных коэффициентов преобразования Фурье высчитывается амплитудно-частотный спектр. На спектре определяются гармоники – характерные «горбы» в районе определенных частот. Амплитуда и частота каждой гармоники определяются внутри окна вычисленного на основе базовой частоты. Базовая частота — это частота пульса, выведенная в Герцах, т.е. ударах в секунду. Она высчитывается следующим образом:

$f_{\text{base}} = 1 / \text{median}(\{\text{RRs}\})$, где $\{\text{RRs}\}$ – множество длин RR-интервалов.

Для каждой гармоники высчитывается частота f_i , амплитуда A_i и энергетический показатель – спектральная плотность E_i .

6. Расчет формы спектра

Формой спектра считается некоторый набор амплитуд исходного спектра, определяемых в рамках исследования как характеристические.

7. Кластеризация выборки кардиограмм

Кластеризация производится для множества кардиограмм, выделенных по определенным признакам. Для заданий кластеризация производится для всех кардиограмм отдельного пациента с различными параметрами кластеризации по выделенным формам спектра.

Значение расстояния от формы спектра до кластера считается как расстояние от формы спектра до средней линии кластера, которая хранится всегда как промежуточный результат. Слияние кластеров сопровождается пересчетом значения средней линии в соответствии с количеством кардиограмм в каждом из объединяемых кластеров.

Также кластеризация может проводиться другими методами: на основе формы, полученной другим алгоритмом, на основе усредненного кардиоцикла и на основе амплитуд или спектральных плотностей гармоник.

8. Расчет показателей тенденции пациента (биологического возраста)

Для определения тенденции изменения показателей состояния пациента между двумя состояниями, выделяются наиболее популярные кластеры наибольшим числом кардиограмм.

На данном этапе выделяется кластер кардиограмм пациентов младшей возрастной группы и кластер кардиограмм пациентов старшей возрастной группы, вне зависимости от пола. Далее будем называть это «молодой» кластер и «старый» кластер соответственно.

Показатель тенденции $t_{ЭКГ-М}$ - это отношение расстояния от текущей формы спектра пациента до средней линии «молодого» кластера к расстоянию до средней линии «старого» кластера.

Для каждого показателя тенденции считается коэффициент доверия $w_{ЭКГ-М}$, рассчитанный на основе правила треугольника. Наиболее показательны состояния пациента, расположенные близко к прямой между «молодым» кластером и «старым» кластером.

Коэффициент доверия принимает значение от 0 до 1. Чем он больше, тем соответственно показатель является более значимым для исследования.

На основе множества пар $\{t_{ЭКГ-М}, w_{ЭКГ-М}\}$ определяются показатели тенденции за последний день, последние 4 дня, неделю (последние 7 дней), месяц (последние 30 дней), и все время наблюдения пациента. Рассчитываются они по формуле математического ожидания. По запросу показатель тенденции может быть рассчитан за произвольный период времени.

Для каждой кардиограммы рассчитываются 6 значений, которые доступны для внешних систем.

9. Заключение

В статье рассматриваются возможности хранения и обработки кардиологической информации для задач математического анализа с помощью облачных технологий.

Основным результатом проведенной работы является реализация обработки и хранения кардиограмм (любого числа отведений), фотоплетизмограмм (пульсовой волны), эхокардиограмм, сейсмокардиограмм для дальнейшего проведения расчётов в облаке и организация доступа к полученным результатам.

10. Литература

- [1] Berezin, A.A. Resonant interaction between the Fermi-Pasta-Ulam recurrences / A.A. Berezin // Bulletin of the Lebedev Physics Institute. FIAN. – 2004. – Vol. 3. – P. 13.
- [2] Andreu, J. Big Data for Health / J. Andreu, C.C.Y. Poon, R.D. Merrifield, G.-Zh. Yang // [Electronic resource]. – Access mode: https://www.researchgate.net/publication/280124446_Big_Data_for_Health.
- [3] Braunwald, E. Heart Disease: A Textbook of Cardiovascular Medicine, Fifth Edition / E. Braunwald. – Philadelphia: W.B. Saunders Co., 1997. – 108 p.
- [4] Disruptive technologies: Advances that will transform life, business, and the global economy / Institute McKinsey [Electronic resource]. – Access mode: <https://www.mckinsey.com/business-functions/digital-mckinsey/our-insights/disruptive-technologies>.
- [5] Fermi, E. Collected Papers of E. Fermi / E. Fermi, J. Pasta, S. Ulam. – 1955. – Vol. 2. – 978 p.
- [6] Ikezi, H. Solitons in plasmas / H. Ikezi // Solitons in Action. Academic Press, 1978. – P. 163-184.

- [7] Lichtenberg, A.J. Regular and Chaotic Dynamics / A.J. Lichtenberg, M.A. Liberman // Applied Mathematical Sciences. – 1992. – Vol. 38.
- [8] Longren, K.E. Solitons in non-linear electric lines / K.E. Longren // Solitons in Action. Academic Press. – 1978. – P. 138-162.
- [9] Moyer, V.A. Screening for coronary heart disease with electrocardiography: U.S. Preventive Services Task Force recommendation statement / V.A. Moyer // Annals of Internal Medicine. – 2002. – P. 157.
- [10] Novopashin, M.A. Forrester's Concept in Modeling Heart Dynamics / M.A. Novopashin, A.V.D. Shmid, A.A. Berezin // IOSR Journal of Computer Engineering. – 2017. – Vol. 19(3). – P. 113-121.

Cloud technologies in problems of mathematical analysis of cardiac information

E.Y. Zimina^{1,2}, M.A. Novopashin², A.V. Shmid^{1,2}

¹National Research University Higher School of Economics, Tikhonov Moscow Institute of Electronics and Mathematics, Tallinskaya street 34, Moscow, Russia, 123458

²EC-Leasing, Warshavskoye Shosse, 125, Moscow, Russia, 117405

Abstract. The article includes the observation of the cloud services and technologies usage. The article contains a review of mathematical analysis of cardiac information using cloud technology, which produces storage, analysis and forecasting on the basis of owned data. In addition, the authors consider the possibility of integrating cloud technologies with external systems.

Keywords: Cloud technologies, Cardiology, Research workflow organization.