

УДК 004.041

Анализ технологий разработки мобильных приложений с использованием больших массивов данных

О.С. Шмелёв*, В.А. Бутаков, Е.М. Гриценко

Сибирский государственный университет науки и технологий имени академика
М. Ф. Решетнева, Красноярск, Российская Федерация

*E-mail: shmelyo.oleg@yandex.ru

Аннотация. В данной работе проведен анализ технологий разработки мобильных приложений с использованием больших массивов данных, также в работе рассмотрены основные методы обработки больших массивов данных.

Ключевые слова: мобильные приложения, разработка приложений, информационные технологии, большие данные, программное обеспечение, методы обработки

Analysis of mobile application development technologies using big data arrays

O.S. Shmelev*, V.A. Butakov, E.M. Gritsenko

Reshetnev Siberian State University of Science and Technology, Krasnoyarsk,
Russian Federation

*E-mail: shmelyo.oleg@yandex.ru

Abstract. In this paper, the analysis of technologies for the development of mobile applications using big data is carried out, and the main methods of processing big data are considered in the work.

Keywords: mobile applications, application development, information technology, big data, software, processing methods

1. Введение

Разработка приложений для мобильных устройств начинается с выбора площадки iOS либо Android. Только уже следующим этапом является - выбор технологии.

На рынке коммерческой разработки программ для телефона существует негласное правило, в согласовании с которым прежде всего ведется разработка версии приложения под iOS и лишь после удачного релиза, разработка программ для телефона на платформу Android. Выше сказанному есть обоснование:

Аудитория устройств iOS в продвинутых странах вплотную приблизилась, либо даже обогнала аудиторию Android. Статистика StatCounter Global Stats указывает, что в Российской Федерации двадцать пять процентов пользователей предпочитают iOS и семьдесят четыре процента Android. Даже в ситуации, если выпуск будет плохим, опасности для репутации компании будут ниже.

Аудитория iOS это более платежеспособная аудитория. Сами устройства стоят дороже и их владельцы больше готовы заплатить за то, либо другое приложение, ежели пользователи Android, которые, обычно предпочитают бесплатные приложения и товары.

Платформа iOS по сравнению с Android различаются существенно большей стабильностью, т.к. существует не так много актуальных версий iOS, под которые проще вести разработку. Например, большая часть пользователей iOS имеют последнюю версию операционной системы, тогда как крайняя версия Android это предпочтение лишь тех, у кого самые новые устройства. Кроме того под Android существует целый большой ряд оболочек и устройств с совсем неожиданными размерами и разрешениями экранов. Разработку своих оболочек Android разрабатывает практически осуществляет разработку фактически каждый уважающий себя бренд, к примеру, Xiaomi, Realme, Huawei, Honor, Samsung, HTC и другие.

В итоге, процесс корректировки багов и разработки приложения для телефона под iOS существенно более понятен и стабилен.

Подытоживая вышеупомянутое, первым шагом рекомендуется разработать и запустить мобильное приложение под iPhone и уже лишь после чего начинать разработку под Android.

Однако, бывают предпосылки, по которым разработка приложения под Android может встать на первое место: Разрабатываемый продукт нацелен на аудиторию,

использующую в основном Android устройства. Это может быть промо приложение, соц. приложение, либо другая программа для масштабного использования. но без серьезного упора на извлечение дохода за подписки.

Если у заказчика есть необходимость заказать разработку программы для телефона в каком-то городе, например Москва, под определенные устройства для решения конкретных потребностей своих сотрудников или пользователей. Например, разработка приложения для заказа такси, в данном случае не обойтись без больших объемов данных [1].

Также в последние годы наблюдается массовый рост объема данных, так как расширяются возможности сбора информации. Общий размер информации, который может сохраняться всеми глобальными тех. средствами, умножился в двое приблизительно все 40 месяцев на протяжении 80-х годов XX века, а с 2012 года и по наше время (2021г.) каждый день генерируется 2,5 экзбайтов ($2,5 \times 10^{18}$) информации.

Значимость Big Data громадна, так как она содержит в себе данные, полученные из соцсетей, машинные данные, записи голоса и видео, структурированные и неструктурированные данные, а также регулярное их сохранение. В приложениях на мобильных устройствах Big data нередко связана с рекламой и сбором информации.

Big Data переводится с английского на русский язык как «Большие данные». Этим термином определяют большие массивы информации, которые невозможно обработать либо всесторонне изучить с помощью обычных способов с внедрением человеческого труда и настольных компьютеров. Особенностью Big Data заключается к тому же, что массив данных с течением времени не прекращает экспоненциально расти, потому для оперативного изучения собранных материалов нужны вычислительные мощности суперкомпьютеров. Следовательно, для обработки Big Data нужны экономные, инноваторские способы обработки информации и предоставления выводов [2].

Цель исследования изучить большие массивы данных, важность, методики анализа больших массивов данных, а также актуальность использования больших массивов данных в разработке мобильных приложений.

2. Технологии и тенденции работы с Big Data

Изначально в совокупность подходов и технологий включались средства массово-параллельной обработки неопределённо структурированных данных, такие как СУБД NoSQL, алгоритмы MapReduce и средства проекта Hadoop. В дальнейшем к

технологиям больших данных стали относиться и иные решения, обеспечивающие подобные по характеристикам возможности по обработке сверхбольших массивов данных, а также некоторые аппаратные средства.

1. MapReduce — модель распределённых параллельных вычислений в компьютерных кластерах, представленная фирмой Google. Согласно данной модели, приложение делится на большое количество схожих элементарных заданий, производимых на узлах кластера и затем естественным образом сводимых в конечный результат.
2. NoSQL (от англ. Not Only SQL, не только SQL) — общий термин для различных нереляционных баз данных и хранилищ, не обозначает какую-нибудь 1 определенную технологию или же продукт. Обычные реляционные базы данных отлично подходят для довольно быстрых и однотипных запросов, а на трудных и гибко построенных запросах, свойственных для больших данных, нагрузка выше разумных пределов и внедрение СУБД становится неэффективным.
3. Hadoop — свободно распространяемый комплект утилит, библиотек и фреймворк для разработки и выполнения распределённых программ, работающих на кластерах из сотен и тыс. узлов. Является одной из основополагающих технологий больших данных.
4. R — язык программирования для статистической обработки данных и работы с графикой. Обширно применяется для анализа данных и практически стал стандартом для статистических программ.
5. Аппаратные решения. Компании Teradata, EMC и др. предлагают аппаратно-программные комплексы, предназначенные для обработки больших данных. Эти комплексы поставляются как готовые к установке телекоммуникационные шкафы, имеющие кластер серверов и управляющее программное обеспечение для массово-параллельной обработки. Сюда еще также относят аппаратные решения для аналитической обработки в оперативной памяти, в частности, аппаратно-программные комплексы Hana фирмы SAP и комплекс Exalytics фирмы Oracle, не обращая внимания на то, что эта обработка изначально не считается массово-параллельной, а объёмы оперативной памяти 1-го узла ограничиваются несколькими терабайтами [3].

3. Методики анализа больших данных

Есть большое количество всевозможных способов анализа массивов данных, в базе коих лежит инструментарий, взятый из статистики и информатики (например, машинное обучение).

1. A/B testing. Метод, в котором контрольная подборка попеременно сравнивается с другими. Тем самым получается обнаружить лучшую комбинацию характеристик для достижения, к примеру, лучшей ответной реакции покупателей на рекламное предложение. Big data дают возможность выполнить большое количество итераций и таким образом получить статистически достоверный результат.
2. Методы класса Data Mining (добыча данных, интеллектуальный анализ данных, глубинный анализ данных) — совокупность способов обнаружения в данных ранее неизвестных, нетривиальных, практически полезных знаний, важных для принятия решений. К этим способам, в частности, относятся обучение ассоциативным правилам (association rule learning), классификация (разбиение на категории), кластерный анализ, регрессионный анализ, обнаружение и анализ отклонений и др.
 - а) Обучение ассоциативным правилам. Набор методик для выявления взаимосвязей, т.е. ассоциативных правил, между переменными величинами в больших массивах данных.
 - б) Классификация. Набор методик, которые позволяют предсказать поведение потребителей в конкретном сегменте рынка (принятие заключений о приобретении, отток, объем потребления и проч.).
 - в) Кластерный анализ. Статистический метод систематизации объектов по группам за счет выявления наперед не известных общих признаков.
 - г) Регрессионный анализ. Набор статистических методов для выявления закономерности между изменением зависимой переменной и одной или же несколькими автономными. Нередко используется для прогнозирования и предсказаний.
3. Краудсорсинг. Методика сбора данных из большого числа источников силами широкого, неопределённого круга лиц, выполняющих данную работу без введения в трудовые отношения.
4. Смешение и интеграция данных (Data fusion and data integration). Набор техник, позволяющих интегрировать разнородные данные из всевозможных источников

- с целью проведения глубинного анализа (например, цифровая обработка сигналов, обработка естественного языка, охватывая тональный анализ, и др.).
5. Ensemble learning. В данном методе задействуется большое количество предикативных моделей за счет чего увеличивается качество проведенных прогнозов.
 6. Искусственные нейронные сети, сетевой анализ, оптимизация, в том числе генетические алгоритмы (genetic algorithm — эвристические алгоритмы поиска, применяемые для решения задач оптимизации и моделирования путём случайного подбора, комбинирования и вариации искомым характеристик с применением механизмов, аналогичных естественному отбору в природе) В данной методике возможные решения представляют в облике `хромосом`, которые могут комбинироваться и мутировать. Как и в процессе естественной эволюции, выживает наиболее приспособленная особь.
 7. Машинное обучение (Machine learning). Направление в информатике которое преследует цель создания алгоритмов самообучения на базе анализа эмпирических данных.
 8. Natural language processing (NLP). Набор взятых из информатики и лингвистики методик распознавания естественного языка человека.
 9. Сетевой анализ (Network analysis). Набор методик анализа связей между узлами в сетях. Применительно к социальным сетям позволяет анализировать связи между отдельными юзерами, компаниями, сообществами и т.п.
 10. Оптимизация (Optimization). Набор численных методов для редизайна сложных систем и процессов для совершенствования 1-го или нескольких характеристик. Методика может помочь в принятии стратегических решений, к примеру, состава выводимой на рынок продуктовой линейки, проведении инвестиционного анализа и проч.
 11. Распознавание образов (Pattern recognition). Набор методик с элементами самообучения для предсказания поведенческой модели потребителей.
 12. Прогнозное моделирование (Predictive modeling). Набор методик, которые дают возможность воссоздать математическую модель наперед заданного возможного сценария развития событий.
 13. Анализ настроений (Sentiment analysis). В основе методик оценки настроений потребителей лежат технологии распознавания естественного языка человека.

Они дают возможность вычлениить из общего информационного потока сообщения, связанные с интересующим предметом (например, потребительским продуктом). Далее оценить полярность суждения (позитивное или же негативное), уровень эмоциональности и проч.

14. 14 Обработка сигналов (Signal processing). Взятый из радиотехники набор методик, который преследует задачу распознавания сигнала на фоне шума и его последующего анализа.
15. Пространственный анализ (Spatial analysis). Набор в некоторой степени взятых из статистики методик анализа пространственных данных – топологии местности, географических координат, геометрии объектов. Источником больших данных в данном случае нередко выступают геоинформационные системы (ГИС).
16. Статистика (Statistics). Наука о сборе, организации и интерпретации данных, охватывая разработку опросников и проведение опытов. Статистические методы нередко используются для оценочных суждений о взаимосвязях меж теми или иными событиями.
17. Контролируемое обучение (Supervised learning). Набор основанных на технологиях машинного обучения методик, которые дают возможность обнаружить функциональные взаимосвязи в анализируемых массивах данных.
18. Имитационное моделирование (Simulation). Моделирование поведения сложных систем нередко применяется для прогнозирования, предсказания и проработки всевозможных сценариев при планировании. Это метод, позволяющий воссоздавать модели, описывающие процессы так, как они протекали бы в реальности. Имитационное моделирование возможно рассматривать как разновидность экспериментальных испытаний.
19. Анализ временных рядов (Time series analysis). Набор заимствованных из статистики и цифровой обработки сигналов методов анализа циклических последовательностей данных. Одни из явных применений – отслеживание рынка ценных бумаг или заболеваемости пациентов.
20. Неконтролируемое обучение (Unsupervised learning). Набор основанных на технологиях машинного обучения методик, которые дают возможность обнаружить сокрытые функциональные взаимосвязи в анализируемых массивах данных. Содержит схожие черты с Cluster Analysis.

21. Визуализация. Методы графического представления результатов анализа больших данных в виде диаграмм или анимированных изображений для упрощения интерпретации облегчения понимания полученных результатов [4].

4. Заключение

Рынок мобильных приложений очень динамичен и постоянно меняется с новыми тенденциями, которые часто выходят за горизонт. Вот почему, чтобы мобильное приложение успешно развивалось в изменчивой среде, нужно быть в курсе всех последних тенденций с помощью оценки данных в реальном времени.

Решения для больших данных предлагают данные в реальном времени, которые имеют все шансы быть применены предпринимателями для принятия быстрых решений и настройки собственных приложений в согласовании с меняющимися рыночными сценариями и потребностями. Это сделает приложение более актуальным и способным собирать более высокую прибыль для бренда в будущем.

Также технология больших данных помогает разработчикам понять объем трафика, создаваемого их приложениями. Это облегчает им оценку каждого отдельного сегмента для отслеживания вовлеченности и обнаружения сбоев, препятствующих плавной работе.

Они могут легко решить эти проблемы и повысить общую производительность своих мобильных приложений, чтобы свести к минимуму частоту отказов пользователей. Кроме того, клиенты будут более склонны использовать такие приложения, которые поддерживаются страстными программистами, которые всегда стремятся обновить функции и функции, чтобы сохранить функциональность своих приложений [5].

Список литературы

1. Обзор платформ и технологий разработки мобильных приложений: сайт. – 2020. – URL: https://integrator.digital/blog/mobile_apps/razrabotka_mobilnogo_prilozheniya/ (дата обращения: 06.10.2021).
2. Что такое Big data: собрали всё самое важное о больших данных: сайт. – 2021. – URL: <https://rb.ru/howto/chto-takoe-big-data/> (дата обращения 06.10.2021).
3. Big data, методы, техники анализа больших данных: сайт. – 2018. – URL: <https://clck.ru/MSmfz> (дата обращения 06.10.2021).

4. TAdviser – портал выбора технологий и поставщика. Большие данные.: сайт. – 2021. – URL: <https://clck.ru/UrvUE> (дата обращения 06.10.2021).
5. Как БОЛЬШИЕ ДАННЫЕ могут привести к передовой разработке мобильных приложений?: сайт. – 2019. – URL: <https://www.machinelearningmastery.ru/how-big-data-can-result-in-advanced-mobile-app-development-cb0f4cc87560/> (дата обращения 29.11.2021).