

Intel демократизирует использование AI

На состоявшейся конференции Intel Artificial Intelligence Day (Intel AI, 16–17 ноября 2016 г., Сан-Франциско, США) Intel анонсировала свою стратегию поддержки технологий искусственного интеллекта и машинного обучения с целью повышения их производительности при выполнении задач на процессорах Intel, а также упрощения и демократизации их использования во всех сферах бизнеса, повседневной жизни и социальных сферах. Это самая масштабная программа, объявленная Intel для поддержки задач AI, конкретные результаты которой появятся уже в 2017 г.



Иван Кузьмин — руководитель проекта Intel® Data Analytics Acceleration Library, корпорация Intel в странах СНГ.

Введение

Объявления, сделанные на Intel AI в ноябре 2016 г. в поддержку развития технологий искусственного интеллекта, являются самыми масштабными за последнее время. Все состоявшиеся анонсы призваны уменьшить постоянно возрастающую сложность алгоритмов, связанных с развитием технологий AI, а также увеличить их производительность при обработке данных постоянно увеличивающихся в объемах и своем многообразии.

На конференции было сообщено, что корпорация Intel заключила рамочное соглашение о покупке компании Nervana Systems (в рамках сделки предстоит получить одобрение регулирующих органов и согласовать окончательные детали соглашения), признанно-го лидера в этой сфере глубокого обучения. Nervana была основана в 2014 г. (штаб-квартира в Сан-Диего, Калифорния) и предоставляет программное и аппаратное обеспечение, специально оптимизированное для глубокого обучения. Как сказала Диана Брайант, исполнительный вице-президент и руководитель подразделения Data Center Group корпорации Intel: «Интеллектуальная собственность компании и её опыт в области ускорения алгоритмов глубокого обучения помогут укрепить позиции Intel в сфере искусственного интеллекта. Мы планируем использовать опыт Nervana по разработке программного обеспечения для дальнейшей оптимизации библиотеки Intel Math Kernel Library и её последующей интеграции в состав отраслевых стандартов. Опыт Nervana в разработке чипов, ее новейший продукт Nervana Engine помогут усилить портфолио Intel в сфере ИИ, снизить стоимость и повысить производительность процессоров Intel Xeon и Intel Xeon Phi в приложении к задачам глубокого обучения».

Intel также представила подробную информацию о том, где революционная технология от Nervana будет интегрирована в до-

рожную карту ее продуктов. Планируется, что Intel будет тестировать первый кремний (под кодовым названием "Lake Crest") в первой половине 2017 г. и сделает его доступным для ключевых клиентов в конце 2017 г. Кроме того, Intel анонсировала новый продукт процессоров Intel Xeon (под кодовым названием "Knights Crest"). Эти процессоры будут специально оптимизированы для нейронных сетей для обеспечения максимальной производительности при использовании технологий глубокого обучения, а также будут предлагать беспрецедентную вычислительную плотность с высокой пропускной способностью интерконнекта.

«Мы ожидаем, что платформа Intel Nervana будет поддерживать рекордную производительность и значительное сокращение времени при работе со сложными нейронными сетями. До конца этого десятилетия, Intel планирует 100-кратное увеличение производительности при работе алгоритмов глубокого обучения с использованием нейронных сетей», — сказала Диана Брайант.

Она также сообщила, что Intel ожидает, что следующее поколение процессоров Intel Xeon Phi (под кодовым названием "Knights Mill") будет поддерживать до 4-х раз более высокую производительность (более полная информация о производительности и тестировании на www.intel.com/benchmarks, прим. ред.), чем предыдущее поколение процессоров для глубокого изучения, и будет доступно в 2017 г. Кроме того, Intel объявила, что начинает поставлять предварительную версию следующего поколения процессоров Intel Xeon (под кодовым названием "Skylake") для поставщиков облачных услуг. Эти процессоры Intel Xeon, дополненные интегрированным расширенным акселератором AVX-512, позволяют существенно повысить скорость формулирования логических выводов на задачах машинного обучения. Дополнительные возможности и конфигурации будут доступны, когда семейство процессоров будет официально анонсировано в середине 2017 г.

Помимо процессорных технологий, Intel заявила, что будет развивать AI-решения на базе Intel Saffron Technology™ (Intel купила Saffron Technology в октябре 2015 года; Saffron является одной из ведущих платформ для когнитивных вычислений, прим. ред.). Платформа Saffron Technology использует in-totory техники рассуждения и прозрачного анализа гетерогенных данных. Эта технология также хорошо подходит для небольших устройств, делая возможным развертывание интеллектуального анализа в интеграции с IoT и, соответственно, извлекая все преимущества из этого.

Чтобы упростить развертывание, Intel также предоставляет общие, интеллектуаль-

ные API-интерфейсы, которые делают возможным повсеместное (от конечных устройств до облачных решений) таких технологий, как: Intel® RealSense™ камеры и Movidius блоки обработки видео (Movidius vision processing units, VPUs).

Также Intel и Google объявили о создании стратегического альянса, чтобы упростить использование открытой, гибкой и безопасной мультислойной инфраструктуры для бизнеса. Сотрудничество включает в себя технологическую интеграцию, ориентированную на Kubernetes (контейнеры), машинное обучение, безопасности и IoT.

Для проведения дальнейших исследований и разработки стратегии AI, Intel объявила о создании совета Intel Nervana AI, в котором примут участие ведущие отраслевые и академические лидеры. Чтобы сделать AI действительно доступной, Intel представила доступ к своим технологиям и инструментам широкому кругу разработчиков. В дополнение, Intel также представила компилятор Intel Nervana Graph Compiler, чтобы ускорить использование технологии deep learning на процессорах Intel.

AI качественно изменит жизнь общества

Искусственный интеллект (Artificial Intelligence, AI) — это наука и семейство технологий, позволяющих машинам приближаться к возможностям человека: чувствовать, разумно действовать и адаптироваться к реальному миру на основе накопленного опыта и данных, т.е. саморазвиваться.

Машинное обучение (Machine Learning, ML) — обширный подраздел искусственного интеллекта, математическая дисциплина, использующая разделы математической статистики, численных методов оптимизации, теории вероятностей, дискретного анализа, и извлекающая знания из данных (<https://ru.wikipedia.org/>). Различают два типа обучения: 1) обучение по

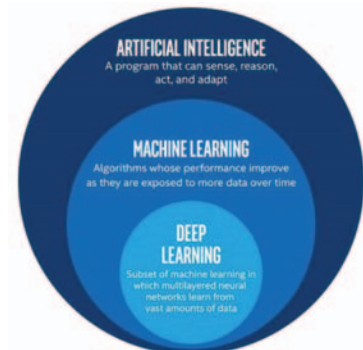


Рис. 1. Таксономия и подчиненность понятий искусственного интеллекта (Artificial Intelligence, AI), машинного обучения (Machine Learning, ML) и глубокого обучения (Deep Learning, DL).

прецедентам, или индуктивное обучение, основано на выявлении закономерностей в эмпирических данных; 2) дедуктивное обучение предполагает формализацию знаний экспертов и их перенос в компьютер в виде базы знаний.

Глубокое обучение (Deep Learning, DL) — часть технологий/методов ML, которая использует модели на базе нейронных сетей для понимания больших объемов данных, позволяя ускорить такие процессы как распознавание изображений, речи, естественного языка и др. (рис. 1)

Одной из основных причин резко возросшего интереса к технологиям ИИ является быстрое развитие микроэлектроники, позволившее за последние 20 лет:

- в десятки раз поднять производительность процессоров на операциях общего назначения и на порядки при обработке отдельных задач с использованием специализированных ускорителей;
- в десятки раз увеличить число ядер в процессоре;
- в десятки раз снизить стоимость за единицу вычислительной мощности и потребление эл/энергии;
- в тысячи раз снизить стоимость за единицу хранения данных и задержки при доступе к данным, одновременно в десятки раз увеличить производительность ввода/вывода;
- в тысячи раз увеличить производительность интерконнекта и широкополосных каналов связи. Завершение разработки стандартов мобильной связи — 5G (Huawei и DOCOMO уже провели первое в истории крупномасштабное эксплуатационное испытание технологий 5G в диапазоне 4,5 ГГц; первые коммерческие сети должны появиться в конце 2019 г.) в корне преобразят нашу жизнь и будут способствовать формированию умного и подключенного общества за счет развития умных городов, беспилотных автомобилей и обеспечения нового уровня промышленной эффективности.

Эти изменения одновременно способствовали и совершенствованию алгоритмов и методов ИИ, что в целом позволило приблизить решение задач ИИ к реальному времени и перевести их из области чисто научного знания в практическую плоскость, одновременно сделав доступность их результатов (в реальном или почти реальном времени) гораздо большему числу потребителей (и обывателей).

Последние объявления Intel не только увеличивают скорость обработки задач ИИ, но и упрощают, и ускоряют создание алгоритмов ИИ для более широкого круга применений, снимая технологические барьеры на пути использования ИИ и делая результаты ИИ доступными для тех, кто в них нуждается.

В этом контексте Intel трансформируется из компании, ориентированной на рынок ПК, в поставщика технологий для облака и для миллиардов умных, подключенных к сети, устройств.

По прогнозам¹⁾, к 2020 году более 50 млрд машин и устройств будут подключены к интернету и соединены друг с другом,

генерируя данные на несколько порядков больше, чем сегодня. Но данные сами по себе имеют ограниченную ценность. Для того чтобы они были полезны и приносили пользу, с точки зрения развития и гармонизации общества, нужно уметь их анализировать. Именно этим целям служат ИИ-технологии и машинного обучения. Их использование создает условия для качественного расширения возможностей машин/информационных технологий, а также применения ИТ. В результате, машины приобретают свойства ощущать, рассуждать, действовать и адаптироваться к реальному миру, а также учиться на основе прошлого опыта и накопленных данных и знаний. Искусственный интеллект увеличивает человеческие возможности, позволяя автоматизировать трудоемкие или опасные виды деятельности человека, делать научные открытия и решать сложные социальные проблемы.

Однако не надо думать, что использование ИИ — задача далекого будущего. Отдельные элементы и технологии ИИ уже активно используются в нашей жизни, например, распознавание голоса и отдельных объектов/лиц в видеопотоке по заданным признакам, персонализированное лечение заболеваний с учетом особенностей генома, противодействие угрозам информационной безопасности, автономные автомобили/дроны (без непосредственного управления оператором) и др.

Охватывая такие вычислительные методы и технологии как расширенная аналитика данных, компьютерное зрение, обработка естественного языка и машинное обучение, ИИ становится одной из самых захватывающих и волнующих областей развития человеческого знания, потому что дает более глубокое понимание сложных данных, что, в свою очередь, является ключом к получению новых ценностей и сервисов для бизнеса и общества в целом. В качестве примеров использования ИИ можно привести следующие: возможность более раннего и точного диагностирования причин заболеваний, что способствует сокращению сроков и стоимости лечения; возможность сельхозпроизводителями/фермерами увеличения урожайности за счет отслеживания в реальном времени особенностей погоды, состояния почвы и др.

Тенденции, способствующие использованию AI

По оценкам²⁾, глобальный рынок робототехники и AI вырастет до \$153 млрд к 2020 г., в котором доля робототехники будет состав-

- 1) "The Internet of Things: How the Next Evolution of the Internet Is Changing Everything." Cisco, April 2011, [cisco.com/c/dam/en_us/about/act79/docs/innov/iot_IBSG_0411FINAL.pdf](http://www.cisco.com/c/dam/en_us/about/act79/docs/innov/iot_IBSG_0411FINAL.pdf).
- 2) "Robot Revolution — Global Robot & AI Primer." Bank of America Merrill Lynch, Dec. 2015, [bofam.com/content/dam/bofamimages/documents/PDFs/robotics_and_ai_condensed_primer.pdf](http://www.bofam.com/content/dam/bofamimages/documents/PDFs/robotics_and_ai_condensed_primer.pdf).
- 3) "It'll Be Back: The Return of Artificial Intelligence." Bloomberg Technology, Feb. 2015, <http://www.bloomberg.com/news/articles/2015-02-03/it-ll-be-back-the-return-of-artificial-intelligence>.
- 4) "From \$600 M to \$6 Billion, Artificial Intelligence Systems Poised for Dramatic Market Expansion in Healthcare." Frost & Sullivan, Jan. 2016, [frost.com/newsroom/press-releases/600-m-6-billion-artificial-intelligence-systems-poised-dramatic-market-expansion-healthcare/](http://www.frost.com/newsroom/press-releases/600-m-6-billion-artificial-intelligence-systems-poised-dramatic-market-expansion-healthcare/).
- 5) "The Artificial Intelligence in Google's Self-Driving Cars Now Qualifies as Legal Driver." Fortune, Feb. 2016, fortune.com/2016/02/10/google-self-driving-cars-artificial-intelligence/.
- 6) "Gartner Reveals Top Predictions for IT Organizations and Users for 2016 and Beyond." Gartner, Oct. 2015, [gartner.com/newsroom/id/3143718](http://www.gartner.com/newsroom/id/3143718).

лять \$83 млрд, а доля AI составит \$70 млрд. По тем же оценкам, использование роботов и искусственного интеллекта сможет повысить производительность труда на 30% во многих отраслях промышленности, при одновременном сокращении затрат на рабочую силу в обрабатывающей промышленности на 33%.

Автомобильная и транспортная промышленность будут одними из первых, где получат немедленные преимущества от умных подключенных технологий с глобальной вырубкой развертывания IoT в транспортном секторе в \$325 млрд в 2018 г. (*Roundup of Internet of Things Forecasts and Market Estimates, 2015.* "Forbes (citing IDC), Dec. 2015, forbes.com/sites/louiscolombus/2015/12/27/roundup-of-internet-of-things-forecasts-and-market-estimates-2015/#5578194048a0).

В 2014 году инвесторы вложили более \$300 млн в стартапы AI, что в 20 раз больше в сравнении с \$15 млн в 2010 году³⁾.

По оценкам⁴⁾, рынок систем искусственного интеллекта в области здравоохранения, вырастет с \$633 млн в 2014 г. до более чем \$6 млрд в 2021 г., что в соответствии с теми же исследованиями позволит не только улучшить результаты лечения на 30–40%, но и снизить затраты на лечение на целых 50%.

В 2016 году Национальная администрация США по транспорту и безопасности движения (U.S. National Highway Transportation and Safety) вынесла решение о том, что AI-система управления вождением автомобиля от Google может считаться законным водителем в соответствии с федеральным законом⁵⁾.

К 2020 г. автономное (без участия брокеров/операторов) программное обеспечение будет принимать участие в 5% всех финансовых транзакций⁶⁾.

Инициативы Intel в области ПО по повышению доступности и производительности задач AI

Одна из ключевых целей Intel — сделать технологии ИИ, а, соответственно, и их результаты доступными гораздо более широкому кругу — как потребителей, так и разработчиков прикладных задач, не имеющих специальной глубокой подготовки в области математики и программирования. Другая ключевая задача — повышение производительности алгоритмов машинного обучения при выполнении на процессорах Intel.

Расширенная таксономия технологий и методов ИИ представлена на рис. 2.

Машинное обучение является ключевым инструментом для ИИ, на основе которого осуществляется разработка и применение алгоритмов, а также улучшается их производительность для какой-то задачи на основе опыта или предыдущих итераций.

Для машинного обучения используются два класса методов/алгоритмов установления зависимостей: на основе глубокого обучения, когда весь процесс обучения строится на многоуровневой нейронной сети, и статистических методах, когда используются статистические и другие техники.

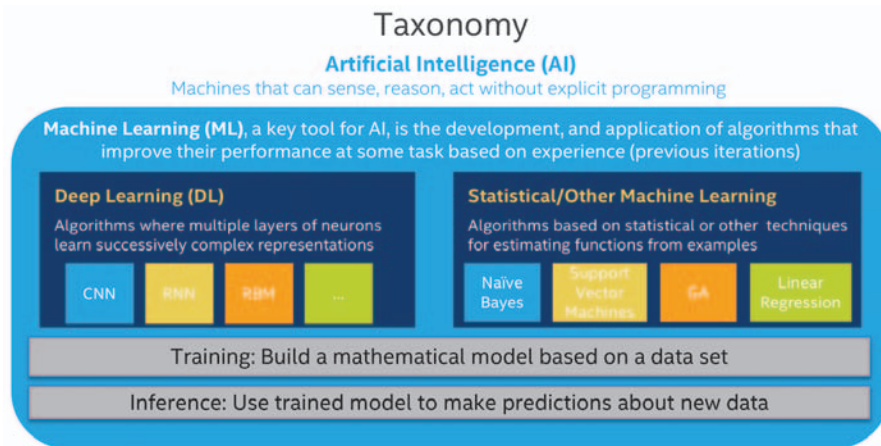


Рис. 2. Расширенная таксономия технологий и методов ИИ.

Основные проблемы, связанные с созданием алгоритмов машинного обучения и их использованием на базе процессоров Intel, следующие:

- фреймворки класса open source для машинного обучения и библиотеки ПО часто не достаточно хорошо оптимизированы для новых систем на базе процессоров Intel;
- существующие фреймворки могут быть трудны в настройке и использовании;
- для существующих фреймворков часто требуется специализированное гетерогенное оборудование – от обучения моделям в центрах обработки данных до развертывания систем на конечных точках.

Чтобы снять все эти проблемы и обеспечить максимальную производительность алгоритмов AI на будущих процессорах, Intel развивает свою всеохватывающую стратегию поддержки технологий AI: от соответствующего инструментария/платформ и оптимизированных фреймворков

до специализированных библиотек и языков программирования с готовыми примитивами (рис. 3). Масштабируемая система Intel Framework предлагает комплексные эталонные архитектуры и решения, которые улучшают технологическую совместимость и снижают сложность развертывания, предлагая пути для широкой адаптации распределенных алгоритмов глубокого изучения и значительного сокращения времени моделирования.

Среди оптимизированных фреймворков для Intel-архитектуры: neon, Spark, Caffe, theano, torch, TensorFlow, dmlc mxnet. В результате этой оптимизации поддерживаются: топологии – AlexNet, GoogLeNet, VGG-19 и др.; однонодовые и мультинодовые конфигурации. Intel-реализации Caffe и Theano уже доступны на GitHub.

Преимущества, которые достигаются в результате реализации стратегии по оптимизации программных кодов для машинного обучения, представлены на рис. 4.

Работая с сообществом open source, Intel оптимизирует промышленные стандарты

для машинного обучения, включая Caffe и Theano, чтобы пользователи могли достигать максимальной производительности, используя технологии Intel, на существующей инфраструктуре. Так, оптимизированная версия Caffe может увеличить производительность до 30 раз по сравнению с основной версией⁸⁾. Intel будет оптимизировать и другие основные фреймворки для машинного обучения для архитектуры Intel.

Intel является одним из ведущих участников нескольких проектов с открытым исходным кодом, в том числе: Linux, Lustre, Embree, OSPRay и OpenSWR. Intel оптимизировала широко используемую библиотеку Intel® Math Kernel Library (Intel® MKL) для создания общедоступных примитивов для алгоритмов машинного обучения, что позволяет иметь более глубокий доступ к оптимизированному коду через стандартный набор API без каких-либо затрат. Кроме того, библиотека Intel® Data Analytics Acceleration Libraries (Intel® DAAL) теперь доступна как open source и поддерживает технологии deep learning.

В 2015 г. Intel на IDF форуме анонсировала сотрудничество с университетом Орегона (Oregon Health & Science University), создав открытое PaaS-решение (Platform-as-a-Service) названное Collaborative Cancer Cloud, которое позволило институтам использовать распределенное машинное обучение для ускорения открытия и изучения новых форм рака и биомаркеров, связанных с развитием рака.

Тестирование оптимизированных примитивов для одного из самых популярных – высокоуровневого языка Python – на базе библиотек Intel® MKL и Intel® DAAL показало увеличение производительности на различных операциях от 10 до 100 раз⁹⁾ (в последнем случае на задачах линейной алгебры) в сравнении с неоптимизированными модулями Python.

Помимо Intel, свои инструменты AI с открытым исходным кодом⁷⁾ выпустили Google, Microsoft, IBM, Facebook и Yahoo.

Заключение

Объявленная Intel программа поддержки задач ИИ и алгоритмов машинного обучения на процессорах Intel позволит быстро создавать и развертывать широкий спектр аппаратных и программных решений, которые являются одновременно открытыми и масштабируемыми, а также имеют повышенную производительность. Эти решения строятся на основе консистентной модели программирования и общей архитектуры, упрощая разработку и использование результатов технологий ИИ во всех сферах жизни и бизнеса.

Иван Кузьмин,
корпорация Intel в странах СНГ



Рис. 3. Компоненты стратегии Intel по поддержке методов и алгоритмов машинного обучения на процессорах Intel.

	Intel® Math Kernel Library (Intel® MKL and MKL-DNN)	Intel® Data Analytics Acceleration Library (Intel® DAAL)	Intel® Distribution for Python*	Intel-Optimized Open Source Frameworks	Intel® Deep Learning SDK	Trusted Analytics Platform
High Level Overview	Высокая производительность математических примитивов допускает низкий уровень контроля	Широкий набор акселераторов объектов для анализа данных специализированной библиотеки поддерживает распределенное машинное обучение на уровне алгоритма	Самые популярный и наиболее быстро развиваемый язык для машинного обучения	Инструментарий, развиваемый наукой и промышленностью, для обучения ML-алгоритмов	Ускорение разработки и развертывания deep learning моделей	Единая платформа от изучения и анализа данных до разработки приложений
Example Usage	Framework developers call matrix multiplication, convolution functions	Call distributed alternating least squares algorithm for a recommendation system	Call scikit-learn k-means function for credit card fraud detection	Script and train a convolution neural network for image recognition	Train and deploy an optimized model on a constrained digital surveillance device	Application creation from the Big Data infrastructure, data science tools up to app development

Рис. 4. Достижимые преимущества при реализации стратегии Intel по созданию специализированных платформ, фреймворков, библиотек для машинного обучения.

7) "Open Source Projects Are Transforming Machine Learning and AI." The Linux Foundation, May 2016, linux.com/news/open-source-projects-are-transforming-machine-learning-and-ai.

8) Up to 30x software optimization improvement claim by Intel is based on customer CNN training workload running 2S Intel® Xeon® processor E5-2680 v3 running Berkeley Vision and Learning Center* (BVL) Caffe + OpenBlas* library and then run tuned on the Intel® Optimized Caffe (internal development version) + Intel® Math Kernel Library (Intel® MKL).

9) Конфигурация при тестировании: Versions: Intel® Distribution for Python 2017 Beta, gcc 15.0; Hardware: Intel® Xeon® CPU E5-2698 v3 @ 2.30GHz (2 sockets, 16 cores each, HT=OFF), 64 GB of RAM, 8 DIMMS of 8GB@2133MHz; Operating System: Ubuntu 14.04 LTS.