

ИСКУССТВЕННЫЙ ИНТЕЛЛЕКТ НА БАЗЕ НЕЙРОСЕТЕЙ ДЛЯ ПРОМЫШЛЕННЫХ ПРЕДПРИЯТИЙ

А.В. Лифанов, Д.И. Соколов (ООО «Сименс»)

Указаны предпосылки появления искусственного интеллекта в промышленности. Рассмотрены аппаратные платформы, реализующие искусственные нейросети. Приведены примеры решений компании «Сименс», использующие алгоритмы искусственных нейросетей, а также примеры реализованных приложений.

Ключевые слова: искусственные нейросети, облачные платформы, обучение нейросети, нейропроцессорный модуль.

Причины появления искусственного интеллекта в производстве

Развитие технологий в промышленности никогда не было более быстрым, чем сегодня. После представления концепции Industry 4.0 на Ганноверской ярмарке в 2013 г. появился целый ряд новых бизнес-моделей, основанных на цифровизации. Парадигма Industry 4.0 подразумевает получение преимуществ от обработки все большего и большего объема производственных данных в реальном времени. Например, для производителя такими преимуществами может стать увеличение прибыли через повышение загрузки имеющегося оборудования, а для потребителя — сроки прихода комплектующих от поставщика для поддержания минимального объема складских запасов. Такая эволюция должна затрагивать всю пирамиду автоматизации, начиная с оборудования в цеху и заканчивая системами принятия решений уровня предприятия и цифрового документооборота между предприятиями.

При этом существуют задачи, которые не могут быть решены обычными средствами автоматизации. К ним относятся процессы, не формализуемые стандартными алгоритмами, требующие знаний эксперта-человека или эксплуатируемые в заранее не полностью известных условиях.

Прообразом искусственного интеллекта в промышленности стали системы нечеткой логики (fuzzy logic) [1]. Идея таких систем возникла в середине 1960-х годов, а первые реализации появились в 1970-х и нашли применение в бытовой технике, фотокамерах, терморегуляторах. При этом вывод выходной величины из входных выполняется по определенным формализованным правилам, которые делают недоопределенные величины пригодными для алгоритмической обработки.

Искусственные нейросети базируются на другом подходе к решению задач управления [2].

Что такое искусственная нейросеть

Искусственная нейросеть строится и работает по принципу нервной системы биологического организма. Она представляет собой систему соединенных и взаимодействующих между собой простых процессоров (нейронов). Каждый такой процессор выполняет над входными параметрами одну несложную математическую операцию (переходную функцию). Соединенная в определенном порядке вся совокупность нейронов позволяет выполнять операции над массивами входящих данных.

Выделяют три этапа построения системы искусственного интеллекта на базе нейросети: выбор архитектуры и подбор первичных параметров; обучение нейросети до требуемых точностных характеристик; дообучение нейросети в процессе работы [2, 3].

Архитектура нейросетей

Выбор архитектуры сети является наиболее сложным этапом. На данный момент разработано несколько десятков стандартизованных архитектур нейросетей, и выбор архитектуры для решения конкретной задачи осуществляется с привлечением экспертов. Алгоритмы автоматизации подбора архитектуры сетей существуют, однако пока больше являются предметом исследований, нежели практики.

Наибольшее прикладное значение имеют три класса архитектур нейросетей [2, 3]:

1) сети прямого распространения (feedforward networks) были разработаны первыми и на данный момент являются наиболее изученными и распространенными. Информация через сеть продвигается напрямую от входов к выходам через промежуточные слои обработки, не имея циклов. Основные области применения таких сетей — классификация объектов и событий, распознавание речи, удаление шума;

2) рекуррентные нейронные сети (recurrent networks) — в отличие от предыдущего класса имеют обратные связи, ведущие от выхода ко входу. Это позволяет реализовать в нейросети функции памяти. Основным назначением таких сетей является прогнозирование и управление, когда имеет значение не только текущее состояние входного массива данных, но и их предыдущая последовательность;

3) сверточные нейросети (convolutional networks) — разновидность сетей прямого распространения, но обычно они образуют не двумерную, а трехмерную архитектуру. Такая архитектура используется в основном при работе со статическими и динамическими изображениями, когда более поверхностные слои сети занимаются поиском и выделением отдельных признаков на изображении, а последующие — группировкой признаков и классификации по ним найденного объекта.

Обучение нейросетей

После выбора архитектуры выполняется обучение нейросети — подбор коэффициентов отдельных нейронов таким образом, чтобы сеть выполняла требуемую задачу с необходимой точностью. Для обучения в большинстве случаев используется выборка в виде пар «вход — правильный выход». Например, таким наборо-

ром для системы распознавания может являться массив фотографий объекта, сделанный в разной ориентации и при разной освещенности. Такой режим называется «обучением с учителем».

Если заранее невозможно подготовить набор правильных значений выходов, возможно обучение нейронов при помощи некой оценочной функции, образующей обратную связь от результата работы предыдущей итерации. Такой режим называется «обучением с подкреплением», и учителем в этом случае выступает неподготовленный набор тестовых данных.

Если подготовка обучающей выборки технически невозможна (изучается поведение людей, поиск аномалий и т.п.), то нейросеть должна самостоятельно сгруппировать поступающие данные в кластеры и выделить квалифицирующие признаки, а соотнесение полученных групп и понятийного поля предметной области необходимо будет выполнить человеку-эксперту [3].

После обучения сети проводится проверка качества ее работы, причем на выборке, которая в обучении не участвовала. Это позволит выявить ошибки в архитектуре или процессе обучения. Известен пример, когда сеть должна была выполнять распознавание танков по фотографиям, но в обучающей выборке они были сделаны на одинаковом фоне. В результате сеть выполняла распознавание не того, что было необходимо, а то, что было проще обобщить (<https://www.porpmech.ru>).

По окончании обучения сеть может использоваться для реальной работы.

Аппаратные платформы для реализации нейросетей

Для реализации алгоритмов искусственной нейросети требуется решить задачу обработки большого массива данных. Для этого вычислительная платформа должна быть оснащена высокоскоростными процессорами и иметь возможность реализовывать параллельные вычисления. Современная микроэлектроника предлагает четыре варианта рабочих платформ для решения таких задач:

- центральные процессоры (CPU);
- графические процессоры (GPU);
- тензорные или нейропроцессоры (TPU, NPU);
- специализированные интегральные схемы (ASIC) [4].

Обычный центральный процессор есть у любого компьютера, но он состоит из небольшого числа ядер. Узким местом при работе с нейросетями может стать производительность интерфейса



Рис. 1. Нейропроцессорный модуль Simatic TM NPU

процессора и памяти. Наилучшую производительность на таких задачах показывают процессоры серии Intel Xeon Phi. Тепловыделение таких чипов достаточно велико, чтобы их можно было применять непосредственно на оборудовании, но они достаточно неплохо показали себя в стоечных промышленных компьютерах.

Современные графические процессоры содержат 3...5 тыс. арифметических блоков, что позволяет успешно решать массивно-параллельные задачи, также у них заметно более скоростной доступ к данным в оперативной памяти. Как следствие, производительность системы на GPU превышает таковую на CPU в 10...100 раз. Графические процессоры наилучшим образом подходят для обучения нейросети, но они менее эффективны в процессе эксплуатации решения на базе нейросети,

так как характеризуются достаточно высоким удельным энергопотреблением. В качестве примера графических ускорителей отметим изделия на базе чипов Nvidia Tesla и Titan, а также ATI FirePro. Платы на базе этих микросхем присутствуют во встраиваемых промышленных компьютерах Simatic IPC.

Понятие тензорного процессора (TPU) было впервые введено компанией Google в 2017 г. На данный момент на рынке присутствует несколько компаний, предлагающих специализированные микросхемы, включающие 2...4 блока матричных вычислений, состыкованные со специализированными высокоскоростными контроллерами памяти. Изначально основным назначением при проектировании таких чипов было ускорение обучения либо ускорение исполнения произвольной нейросети (одно из первых применений — сервис перевода Google Translate) для

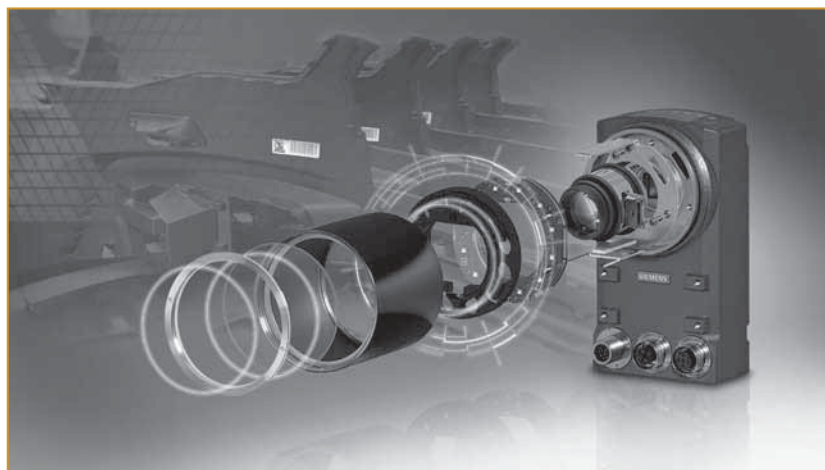


Рис. 2. Интеллектуальная видеочамера Siemens MV 550

дата-центров. Результатом таких разработок является решение, примерно на два порядка более энергоэффективное, чем CPU, и в 10...30 раз более быстрое, чем GPU, однако при этом снижается точность вычислений. По мере развития темы появились решения для компактных компьютеров, что и дало толчок развитию промышленного искусственного интеллекта. Примеры таких чипов — Google TPU, Huawei Ascend, а также Movidius (с 2016 г. — подразделение Intel) Myriad, на базе которого построен модуль Simatic TM NPU (рис. 1).

Внешне Simatic TM NPU представляет собой модуль ввода/вывода для промышленного контроллера серии Simatic S7-1500 и оборудовано разъемами Ethernet и USB 3.1. В это устройство можно загрузить подготовленную и отреннированную нейросеть, которая будет использоваться в качестве источников или приемников информации как непосредственно подключенные устройства, так и входы/выходы контроллера. В настоящее время нейросети для этого модуля разрабатываются в программном обеспечении TensorFlow или Caffe2. В 2020 г. планируется выпуск пакета Siemens AI Toolbox, содержащего набор готовых нейросетей для решения промышленных задач.

Специализированная интегральная схема (ASIC) представляет собой сотни логических блоков, объединенных в единую сеть и находящихся на едином кристалле. При этом помешать задействовать все блоки одновременно может только неравномерность вычислений на разных уровнях сети. ASIC характеризуются хорошими показателями по производительности и энергоэффективности (отношение производительности к энергопотреблению), но крайне тяжело изменить уже залитую в них программу. Таким образом, специализированные интегральные схемы являются отличным решением для исполнения уже разработанных нейросетей при больших тиражах заводского выпуска, например, интеллектуальных видеочамер линейки Siemens MV550 (рис. 2).

Примеры применения нейросетей

Процесс разработки решения на нейросетях на данный момент является достаточно трудоемким и оправдан в тех случаях, когда решение должно иметь гибкость для устойчивого функционирования в заранее неизвестных условиях. В реалиях цифрового производства алгоритмы искусственного интеллекта могут реализовываться в облачном дата-центре или исполняться локально, например, на конкретном производстве.

Решения, исполняемые в дата-центре, обычно направлены на анализ длинных рядов данных, корреляций большого числа параметров, выработку решений в режиме, отличном от реального времени. Одним из требований к реализации данного варианта является наличие хорошего канала связи между местом измерения соответствующих параметров и местом отработки алгоритмов нейросети.

Примером может служить система технического обслуживания и ремонта оборудования на базе анализа телеметрии (ток, скорость, вибрация), предсказание флуктуаций потребления электроэнергии района, поиск нестандартного поведения операторов систем (антифрод). С аппаратной точки зрения такие решения могут быть реализованы как в виде выделенных серверов on-premise, так и в виде арендуемых виртуальных машин или контейнеров операторов облачных решений.

Локальные применения реализуются по идеологии периферийных вычислений (edge computing) и обычно предназначены для оперативного управления. Иллюстрациями таких решений может стать поиск посторонних включений в сыпучих материалах на конвейере, управление захватом робота при неизвестном исходном положении объекта, оценка температуры пламени по его цвету, расчет параметров, которые невозможно померить непосредственно. При реализации данного варианта повышаются требования к вычислительной мощности, установленной в шкафу управления, зато резко снижаются потребности в стабильном и высокоскоростном канале связи.

Применение облачной платформы MindSphere

Для реализации задач обслуживания и ремонта оборудования может успешно применяться платформа MindSphere, объединяющая возможности операционной системы и облачной платформы для интернета вещей, позволяющая подключать любые физические устройства и датчики к цифровому информационному пространству. Среди прочего платформа обладает всеми необходимыми инструментами для построения систем предиктивной диагностики. Так, в пакет услуг MindSphere Predictive Learning входит набор алгоритмов и библиотек для создания предсказательных моделей с помощью методов глубокого машинного обучения, нейронных сетей и на основе математических моделей. В MindSphere готовые аналитические модули могут быть быстро сконфигурированы и подключены к источникам промышленных данных, отображающим работу промышленных систем.

С помощью модуля Data Science Workbench, входящего в пакет Predictive Learning, пользователи могут создавать и адаптировать модели, сохраняя их затем в хранилище Zeppelin Notebook. Каждый экземпляр модели может быть обучен на серии реальных данных, собранных в MindSphere, а результат анализа — визуализирован. Модели могут быть созданы с использованием известных алгоритмов и наборов библиотек, таких как TensorFlow, Spark, NumPy, Scikit-learn, Keras, SciPy, Matplotlib, Pandas, а также библиотек распределенных вычислений, например, MLib. У специалиста по анализу данных также есть возможность конфигурирования необходимых вычислительных ресурсов для работы моделей. Predictive Learning позволяет быстро анализировать и трансформировать большие массивы данных без необходимости программирования или создания скриптов.

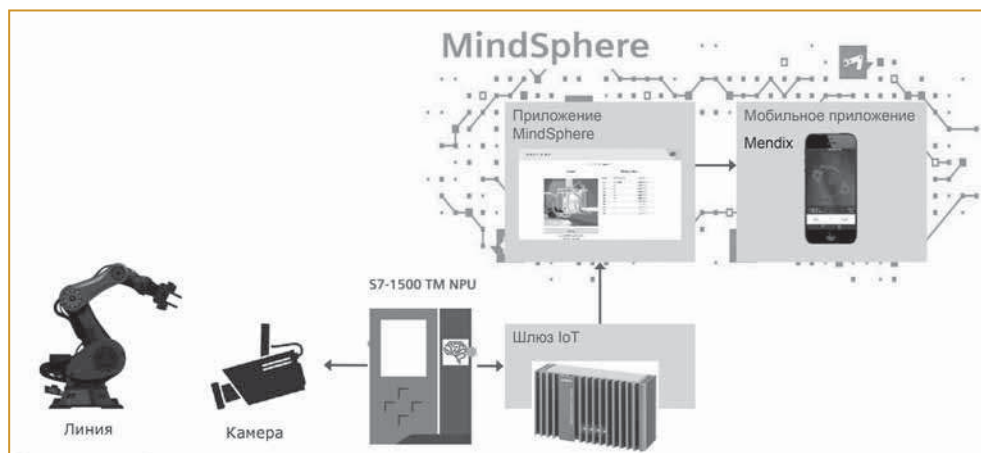


Рис. 3. Структура системы контроля качества

Контроль качества изделий

Производитель OEM-изделий в автоиндустрии столкнулся с проблемой отклонений качества комплектующих от одного из поставщиков. Отклонения обнаруживались только, когда изделие было уже собрано, и влекли за собой расходы по пересборке. Для предотвращения пересборок требовалась информация о качестве изделий на линии поставщика в реальном времени. Для решения задачи была создана гибкая платформа контроля качества на базе платформы MindSphere с использованием Mendix¹.

На выходе производственной линии каждое изделие помечается RFID-меткой и снимается на видеокамеру. Информация с видеокамеры обрабатывается в специализированном аппаратном модуле с помощью алгоритмов искусственного интеллекта. Информация о годных и бракованных комплектующих, а также уровень уверенности в сделанном заключении сохраняются, анализируются и визуализируются в отчетах MindSphere. На базе платформы Mendix разработано мобильное приложение, которое информирует менеджера линии о результатах с низкой уверенностью, при необходимости сеть дообучается. Результаты анализа качества передаются роботу, который сортирует изделия по RFID и качеству (рис. 3).

Систему сортировки комплектующих удалось запустить за несколько недель на конвейере поставщика.

Предиктивная диагностика отказов ветрогенераторов

Примером предиктивной диагностики отказов ветрогенераторов является система Pythia, создан-

ная компанией Siemens Gamesa. Данное решение выполняет анализ данных более чем с 11 тыс. ветрогенераторов, собирая примерно 2300 параметров с каждого. Система анализирует исторические наборы данных и делает вывод о происхождении аномалий в них и о примерном сроке до отказа компонента. Данная информация, с одной стороны, может автоматически планировать работу обслуживающего персонала,

а с другой — является исключительно важным источником информации для конструкторского бюро, разрабатывающего следующие поколения генераторов.

Заключение

Использование технологий искусственного интеллекта в современных промышленных приложениях уже показало свою перспективность. На сегодняшний день заказчикам и интеграторам доступны как оборудование, так и алгоритмы, пригодные для внедрения. Цифровая трансформация промышленных предприятий представляет сегодня широкое поле деятельности, работа на котором может привести к заметному экономическому эффекту. Остается надеяться, что осведомленность руководства и уровень подготовки кадров не станут ограничивающими факторами для внедрения описываемых решений.

Список литературы

1. Norm Dingle. Artificial Intelligence: Fuzzy Logic Explained//Control Engineering. 2011. November.
2. Prakruthi Gowd B, Manjunath N. Hegde. Comparison of Artificial Neural Networks and Fuzzy Logic Approaches for Crack Detection in a Beam Like Structure // IJAIA. 2018. Vol. 9. No.1.
3. Тадеусевич Р., Боровик Б., Гончарж Т., Леннер Б. Элементарное введение в технологию нейронных сетей с примерами программ. Перевод И. Д. Рудинского. М.: Горячая линия — Телеком. 2011. 408 с.
4. Gaurav Batra, Zach Jacobson, Siddarth Madhav, Andrea Queirolo, and Nick Santhanam. Artificial-intelligence hardware: New opportunities for semiconductor companies. McKinsey & Company, 2018.

*Лифанов Александр Витальевич — специалист по продукту,
Соколов Дмитрий Игоревич — эксперт по развитию направления
облачных технологий Управление «Цифровое производство», ООО «Сименс».
Контактный телефон (495) 737-17-37.*

¹ Mendix - это платформа приложений как услуга (aPaaS) для быстрой разработки приложений. Быстрое предложение aPaaS отдает приоритет скорости и гибкости и использует принцип "без кода".