

Системы искусственного интеллекта

«02.03.03 - Математическое обеспечение и администрирование информационных систем
направленность разработка и администрирование информационных систем»

<http://vikchas.ru>

Тема 2. Введение в искусственный интеллект и машинное обучение Лекция 3 «Области искусственного интеллекта»

Часовских Виктор Петрович

д-р техн. наук, профессор кафедры ШИиКМ

ФГБОУ ВО «Уральский государственный экономический
университет»

Екатеринбург 2022

Области искусственного интеллекта

Искусственный интеллект включает в себя множество областей математики и информационных технологий, а также экономики, биологии, физики, социологии и других наук.

В истории искусственного интеллекта мы рассмотрели только самые известные и значимые события и подходы, разумеется, идей создания искусственного интеллекта было гораздо больше.

Например, существует подход на основе эволюционных алгоритмов: он заключается в том, чтобы попытаться имитировать «эволюцию» с помощью случайных «мутаций» программы; сегодня такие подходы используются совместно с современными системами искусственного интеллекта, например с нейронными сетями.

Помимо понимания искусственного интеллекта как способности компьютера решать интеллектуальные задачи подобно человеку, существует понимание искусственного интеллекта как создания компьютера, имитирующего **человеческий мозг**.

Однако в науке пока нет полного понимания, как работает мозг, поэтому способов его искусственного повторения тоже **не существует**.

На сегодняшний день технологии искусственного интеллекта и обработки больших объемов данных активно используются в бизнесе.

Будем выделять несколько смежных областей, отвечающих за разработку этих технологий.



Машинное обучение

Машинное обучение (machine learning, ML) - это одно из направлений разработки ИИ, основанное на выполнении компьютером множества сходных задач без использования прямых инструкций.

Машинное обучение базируется на трех основных понятиях:

- **алгоритмы** - специальные программы, «подсказывающие» компьютеру, каким источником данных необходимо воспользоваться. Для каждой задачи подбираются отдельные алгоритмы, составленные с расчетом на ускорение обработки данных и получение точного результата;
- **наборы данных** - информация (выборки данных) в виде текстовых, графических, видеофайлов, которую машина использует для накопления опыта при обучении. При этом для решения каждого конкретного типа задач в систему должны загружаться уникальные данные;
- **признаки** (свойства, метрики, фичи, features) - индивидуальные измеримые параметры наблюдаемых явлений, от правильности подбора которых зависит успешность и скорость machine learning.

Оптимальными для правильного проведения машинного обучения считаются выборки данных, составленные вручную и содержащие максимум информации разного качества - это позволяет компьютеру выявлять неочевидные взаимосвязи между данными и делать по ним полезные выводы.

Машинное обучение используется для решения задач по категориям:

регрессии - составления прогнозов на основе выборки данных с отличающимися признаками;

классификации - получения конкретного ответа на основании набора признаков;

кластеризации - разбивки данных на несколько групп;

уменьшения размерности - сокращения большого количества признаков для удобства их дальнейшей визуализации;

выявления аномалий - поиска отличий в наборах данных от стандартной информации.

В зависимости от того, по какому принципу осуществляется машинное обучение, оно делится на:

обучение с учителем, которое предполагает использование полного набора снабженных признаками данных (размеченного дата сета) для тренировки системы ИИ на всех стадиях ее создания;

обучение без учителя, при котором машина самостоятельно выявляет закономерности, определяет признаки и классифицирует полученные данные;

обучение с подкреплением - когда машине требуется правильно решить поставленные перед ней задачи во внешней среде, располагая несколькими возможными вариантами действия.

Для создания программ машинного обучения используются различные языки.

Приведем список на базе индекса TIOBE языков программирования искусственного интеллекта и машинного обучения.

Индекс TIOBE — это обновляемый раз в месяц показатель популярности языков программирования. Рейтинги основаны на количестве квалифицированных инженеров по всему миру, курсах и сторонних поставщиках. Для расчёта используются популярные поисковые системы, такие как Google, Bing, Yahoo!, Wikipedia, Amazon, YouTube и Baidu. Важно отметить, что этот индекс — это не рейтинг лучших языков программирования или языков с большим числом строк кода.

| Январь 2022 г. | Январь 2021 г. | Изменять | Язык программирования | Рейтинги | Изменять |
|----------------|----------------|----------|----------------------------------------------------------------------------------------------------|----------|----------|
| 1 | 3 | ▲ |  ПИТОН | 13,58% | +1,86% |
| 2 | 1 | ▼ |  C | 12,44% | -4,94% |
| 3 | 2 | ▼ |  Ява | 10,66% | -1,30% |
| 4 | 4 | |  C++ | 8,29% | +0,73% |
| 5 | 5 | |  C# | 5,68% | +1,73% |
| 6 | 6 | |  Visual Basic | 4,74% | +0,90% |

| | | | | | | |
|----|----|----|-------------------------------------------------------------------------------------|----------------------|-------|--------|
| 7 | 7 | |  | JavaScript | 2,09% | -0,11% |
| 8 | 11 | ↑ |  | язык ассемблера | 1,85% | +0,21% |
| 9 | 12 | ↑ |  | SQL | 1,80% | +0,19% |
| 10 | 13 | ↑ |  | Быстрый | 1,41% | -0,02% |
| 11 | 8 | ↓ |  | PHP | 1,40% | -0,60% |
| 12 | 9 | ↓ |  | р | 1,25% | -0,65% |
| 13 | 14 | ↑ |  | Идти | 1,04% | -0,37% |
| 14 | 19 | ↑↑ |  | Делфи/Объект Паскаль | 0,99% | +0,20% |

| | | | | | | |
|----|----|----|------------------------------------------------------------------------------------|---------------------------|--------------|---------------|
| 15 | 20 | ⬆️ |  | Классический Visual Basic | 0,98% | +0,19% |
| 16 | 16 | |  | МАТЛАБ | 0,96% | -0,19% |
| 17 | 10 | ⬇️ |  | заводной | 0,94% | -0,90% |
| 18 | 15 | ⬇️ |  | Рубин | 0,88% | -0,43% |
| 19 | 30 | ⬆️ |  | Фортран | 0,77% | +0,31% |
| 20 | 17 | ⬇️ |  | Перл | 0,71% | -0,31% |

В качестве среды разработки используются платформы и среды Visual Studio 22, R-Studio, R-Brain, Eclipse, PyCharm, Spyder, IntelliJ IDEA, Jupyter Notebooks, Juno.

После создания [SQL Server 2019](#), ориентированного на использование технологии больших данных и машинного обучения в среде платформы [ASP.NET Core MVC 5](#) следует ориентироваться на [Visual Studio 22](#).

Глубинное обучение

Иногда называют «глубокое обучение» (от англ. Deep learning). Подобласть машинного обучения, где в качестве алгоритмов используются нейронные сети.

форма машинного обучения, которая предусматривает извлечение, или моделирование, признаков данных с использованием сложных многослойных фильтров. Поскольку глубинное обучение является весьма общим способом моделирования, оно способно решать сложные задачи, такие как компьютерное зрение и обработка естественного языка. Этот подход существенно отличается от традиционного программирования, и от других методов машинного обучения.

Глубинное обучение не только может дать результат там, где другие методы не сработают, но и позволяет построить более точную модель или же сократить время на ее создание; расплачиваться же за это приходится еще большими вычислительными мощностями. Еще одним недостатком глубинного обучения является сложность интерпретации получаемых моделей.

Определяющая характеристика глубинного обучения — наличие более одного слоя между входом и выходом. Обычно, говоря о глубинном обучении, подразумевают использование глубоких нейронных сетей. Есть, однако, несколько алгоритмов, которые реализуют глубинное обучение, используя иные типы слоев.

Data Science

Это концепция объединения статистики, анализа данных, машинного обучения и связанных с ними методов для понимания и анализа реальных явлений.

Data Science — это работа с большими данными (англ. Big Data).

Большие данные — это огромные объёмы неструктурированной информации: например, метеоданные за какой-то период, статистика запросов в поисковых системах, результаты спортивных состязаний, базы данных геномов микроорганизмов и многое другое. Ключевые слова здесь — «огромный объём» и «неструктурированность». Чтобы работать с такими данными, используют математическую статистику и методы машинного обучения.

Специалист, который делает такую работу, называется дата-сайентист (или Data Scientist).

Он анализирует большие данные (Big Data), чтобы делать прогнозы.

Какие именно прогнозы — зависит от того, какую задачу нужно решить. Итог работы дата-сайентиста — прогнозная модель. Если упростить, то это программный алгоритм, который находит оптимальное решение поставленной задачи.

Data Science отличается от бизнес-аналитик.

Основная разница заключается в результате. Data Scientist ищет в массивах данных **связи и закономерности**, которые позволят ему создать модель, предсказывающую результат, — то есть можно сказать, что Data Scientist работает на будущее. Он использует программные алгоритмы и математическую статистику и решает поставленную задачу в первую очередь как техническую.

Бизнес-аналитик сосредоточен не столько на технической, программной стороне задачи, сколько на коммерческих показателях компании.

Он работает со статистикой и может оценить, например, насколько эффективна была рекламная кампания, сколько было продаж в предыдущем месяце и так далее.

Вся эта информация может использоваться для улучшения бизнес-показателей компании. Если данных много и нужен какой-то прогноз или оценка, то для решения технической стороны этой задачи бизнес-аналитик может привлечь дата-сайентистов.

Data Mining

Широкое понятие, означающее извлечение **знаний** из **данных**.

Data mining переводится как “добыча” или “раскопка данных”.

Нередко рядом с Data Mining встречаются слова “обнаружение знаний в базах данных” (knowledge discovery in databases) и “интеллектуальный анализ данных”. Их можно считать синонимами Data Mining.

Возникновение всех указанных терминов связано с новым витком в развитии средств и методов обработки данных.

В общем технологию **Data Mining** достаточно точно определяет Григорий Пиатецкий-Шапиро – один из основателей этого направления:

Data Mining – это процесс обнаружения в сырых данных

- ранее неизвестных
- нетривиальных
- практически полезных и доступных интерпретации знаний,
- необходимых для принятия решений в различных сферах человеческой деятельности.

Data Mining является мультидисциплинарной областью, возникшей и развивающейся на базе достижений прикладной статистики, распознавания образов, методов искусственного интеллекта, теории баз данных и др. Отсюда обилие методов и алгоритмов, реализованных в различных действующих системах Data Mining. Многие из таких систем интегрируют в себе сразу несколько подходов. Тем не менее, как правило, в каждой системе имеется какая-то ключевая компонента, на которую делается главная ставка.

Области применения технологии Data Mining



Схема преобразования данных с использованием технологии Data Mining



Большие данные

Это набор подходов и методов, разработанных для анализа данных огромных размеров.

Стоит отметить, что не всякая работа с данными относится к искусственному интеллекту (например, аналитик, делающий вывод по графикам, не относится к искусственному интеллекту), и не все алгоритмы искусственного интеллекта разрабатываются с использованием данных (например, экспертные системы, упомянутые выше).

Тем не менее, множество современных интеллектуальных систем основано именно на обучении по данным: машинный перевод, распознавание изображений и речи, прогнозирование поведения клиентов и др. Во время обучения по данным алгоритм «изучает» большое количество реальных случаев (например, поведения клиентов или переводов текстов) и благодаря этому делает качественные предсказания для новых случаев.

Источники сбора больших данных делятся на три типа:

- социальные;
- машинные;
- транзакционные.

Все, что человек делает в компьютерной сети, — источник **социальных** больших данных. Каждую секунду пользователи загружают в Instagram 1 тыс. фото и отправляют более 3 млн электронных писем. Ежесекундный личный вклад каждого человека — в среднем 1,7 мегабайта.

Другие примеры **социальных** источников Big Data — статистики стран и городов, данные о перемещениях людей, регистрации смертей и рождений и медицинские записи.

Большие данные также генерируются **машинами**, датчиками и «интернетом вещей». Информацию получают от смартфонов, умных колонок, лампочек и систем умного дома, видеокамер на улицах, метеоспутников и т.п..

Транзакционные данные возникают при покупках, переводах денег, поставках товаров и операциях с банкоматами.

Концептуально мы рассмотрели составляющие или технологии искусственного интеллекта

