Нужен ли здесь ML? Критерии ML-системы и стадии ее развития

Page • 1 backlink • uni

Важный вопрос при решении задачи: нужен ли вообще ML? Зачастую эвристики могут без больших потерь денег принести львиную долю возможного качества

Примеры простых задач:

- выявление персональных данных
- обработка 10 запросов в день (нужен ли GPU?)
- определение да/нет в записи голоса

Машинное обучение нужно, когда речь идет о запоминании сложных паттернов и, что не менее важно, умении работать на ранее не виданных данных

Также нужны большие объемы актуальных данных – язык алеутов, на котором говорит около 300 человек, интересен для академии, но не для бизнеса

Также нужно иметь представление о том, что ответы, которые нам нужны, вообще можно предсказать (иначе даже ML не поможет)

Задача с озвучиванием книг: тут автоматизация даст прирост по деньгам, поскольку использовать дикторов еще дороже

Как правило, цель бизнеса вообще не сходится со стандартными метриками ML и метрики бизнеса тяжело формализовать с точки зрения качества моделей (определяем бизнес-эффект → определяем критерии успеха → вводим иерархию ML-метрик, если сразу на этом обучать модели и делать MVP, такое прокатит только у большого бизнеса, имеющего большой опыт в переводе критериев успеха в метрики ML)

Важный критерий: работать надежно при должном уровне качества

Надежность определить тяжело: вместо 500-ок от сервиса получаем галлюцинации, которые быстро не починить (и не отловить тестами) – приходится откатывать и терпеть большие PR-риски

Другой важный критерий: масштабируемость – в разное время нагрузка разная, поэтому нужно уметь подключать нужное количество ресурсов в нужный момент (если использовать все доступные мощности постоянно, деньги не сойдутся)

С другой стороны: у больших сервисов малое время downtime уже катастрофически влияет на деньги, поэтому нужно иметь подушку безопасности

Третий критерий: поддерживаемость – в нашем сервисе должен быть минимально возможный стержень нужных компонент, задачи которых ясны, а функциональность задокументирована

Против масшабируемости и надежности, как правило, стоит эффективность по деньгам

Поэтому перед тем, как бежать обучать модели, нужно определиться с трейд-оффом требований выше

Жизненный цикл ML-системы:

- формулировка образа проекта, целеполагание (бизнес-метрики + технические)
- определиться с набором критериев (например, сколько пользователей)
- методы оценки выполнения заложенного целеполагания и соблюдения набора критериев
- определение объема ресурсов
- временные рамки

При оценке ресурсов хороший пример: трейд-офф между latency и throughput – по мере развития можно обменивать одно на другое, но важно, что оба этих значения считаются и о них важно иметь представление

Работа с данными (DataOps):

- источники данных
- формат данных
- обработка данных
- хранилище
- потребитель данных
- контроллер данных (может быть чисто фиктивным, а может и присутствовать в лице команды ИБ со своими правилами работы с данными и пайплайнами)

В текущих реалиях гораздо выгоднее собрать качественные данные для обучения существующей модели, чем уходить в R&D и совершенствовать архитектуру

От модели важна способность "запомнить данные", а не "мыслить аналогично мозгу" (последнее – больше про науку, а в проде есть та или иная возможность заменить модель, обновить на более акуальную)

Разработка модели:

- создание датасета
- разработка признаков
- обучение моделей
- оффлайн-оценка
- фиксы в проде и DataOps
- поиск необходимых ресурсов – даже если нашли модель, которая будет с нами "до конца", этот блок всегда актуален (также возможно ускорение работы нашей модели)

Деплой:

- деплой и обслуживание (создание сервиса)
- стратегии релиза
- онлайн-оценка
- быстрые фиксы падений

Сам по себе деплой несложен, сложен надежный деплой

По большей части масштабирование – это особенно про большие компании, в них очень важен качественный MLOps для того, чтобы от очередного внедрения бизнес не мог понести большие финансовые потери и PR-риски

Бизнес-аналитика:

- пользовательский опыт
- связывание показателей модели с бизнес-показателями