ВВЕДЕНИЕ: ИСТОРИЯ И СТРУКТУРА АІ И МЬ

Сергей Николенко

СПбГУ — Санкт-Петербург 01 сентября 2017 г.

Random facts:

- 1 сентября 1827 г. в России был введён запрет принимать крестьян в высшие и средние учебные заведения
- 1 сентября 1938 г. в «Правде» был впервые опубликован лозунг «За Родину! За Сталина!»
- 1 сентября 2012 г. в России вступил в силу закон «О защите детей от информации, причиняющей вред их здоровью и развитию»

Первые мысли об искусственном интеллекте

- Гефест создавал себе роботов-андроидов, например, гигантского человекоподобного робота Талоса.
- Пигмалион оживлял Галатею.
- Иегова и Аллах куски глины.
- Особо мудрые раввины могли создавать големов.
- Альберт Великий изготовил искусственную говорящую голову (чем очень расстроил Фому Аквинского).
- Начиная с доктора Франкенштейна, дальше AI в литературе появляется постоянно...

ТЕСТ ТЬЮРИНГА

- АІ как наука начался с теста Тьюринга (1950).
- Компьютер должен успешно выдать себя за человека в (письменном) диалоге между судьёй, человеком и компьютером.
- Правда, исходная формулировка была несколько тоньше и интереснее...

ТЕСТ ТЬЮРИНГА

- Здесь уже очевидно, сколько всего надо, чтобы сделать Al:
 - обработка естественного языка;
 - представление знаний;
 - выводы из полученных знаний;
 - обучение на опыте (собственно machine learning).

Дартмутский семинар

- Термин AI и формулировки основных задач появились в 1956 на семинаре в Дартмуте.
- Его организовали Джон Маккарти (John McCarthy), Марвин Мински (Marvin Minsky), Клод Шеннон (Claude Shennon) и Натаниэль Рочестер (Nathaniel Rochester).
- Это была, наверное, самая амбициозная грантозаявка в истории информатики.

Дартмутский семинар

Мы предлагаем исследование искусственного интеллекта сроком в 2 месяца с участием 10 человек летом 1956 года в Дартмутском колледже, Гановер, Нью-Гемпшир. Исследование основано на предположении, что всякий аспект обучения или любое другое свойство интеллекта может в принципе быть столь точно описано, что машина сможет его симулировать. Мы попытаемся понять, как обучить машины использовать естественные языки, формировать абстракции и концепции, решать задачи, сейчас подвластные только людям, и улучшать самих себя. Мы считаем, что существенное продвижение в одной или более из этих проблем вполне возможно, если специально подобранная группа учёных будет работать над этим в течение лета.

1956-1960: БОЛЬШИЕ НАДЕЖДЫ

- Оптимистическое время. Казалось, что ещё немного, ещё чуть-чуть...
- · Allen Newell, Herbert Simon: Logic Theorist.
 - Программа для логического вывода.
 - Смогла передоказать большую часть *Principia Mathematica*, кое-где даже изящнее, чем сами Рассел с Уайтхедом.

1956-1960: БОЛЬШИЕ НАДЕЖДЫ

- Оптимистическое время. Казалось, что ещё немного, ещё чуть-чуть...
- General Problem Solver программа, которая пыталась думать как человек;
- Много программ, которые умели делать некоторые ограниченные вещи (microworlds):
 - · Analogy (IQ-тесты на «выберите лишнее»);
 - · Student (алгебраические словесные задачи);
 - · Blocks World (переставляла 3D-блоки).

1970-E: KNOWLEDGE-BASED SYSTEMS

- Суть: накопить достаточно большой набор правил и знаний о предметной области, затем делать выводы.
- Первый успех: MYCIN диагностика инфекций крови:
 - · около 450 правил;
 - результаты как у опытного врача и существенно лучше, чем у начинающих врачей.

1980-е: коммерческие применения; индустрия АГ

- Началось внедрение.
- Первый AI-отдел был в компании DEC (Digital Equipment Corporation);
- \cdot Утверждают, что к 1986 году он экономил DEC \$10 млн. в год;
- Бум закончился к концу 80-х, когда многие компании не смогли оправдать завышенных ожиданий.

1990-2010: DATA MINING, MACHINE LEARNING

- В последние десятилетия основной акцент сместился на машинное обучение и поиск закономерностей в данных.
- Особенно с развитием интернета.
- Сейчас про AI в смысле трёх законов робототехники уже не очень вспоминают.
- \cdot // Но роботика процветает и пользуется machine learning на каждом шагу.

Определение

• Что значит — обучающаяся машина? Как определить «обучаемость»?

Определение

• Что значит — обучающаяся машина? Как определить «обучаемость»?

Definition

Компьютерная программа обучается по мере накопления опыта относительно некоторого класса задач T и целевой функции P, если качество решения этих задач (относительно P) улучшается с получением нового опыта.

- Определение очень (слишком?) общее.
- Какие конкретные примеры можно привести?

ЧЕМ МЫ БУДЕМ ЗАНИМАТЬСЯ

- Мы будем рассматривать разные алгоритмы, которые решают ту ли иную задачу, причём решают тем лучше, чем больше начальных (тестовых) данных ему дадут.
- Сегодня мы поговорим об общей теории байесовского вывода, в которую обычно можно погрузить любой алгоритм машинного обучения.
- Но сначала краткий обзор основных задач машинного обучения в целом.

- Обучение с учителем (supervised learning) обучение, в котором есть некоторое число примеров с правильными ответами:
 - обучающая выборка (training set) набор примеров, каждый из которых состоит из признаков (features, attributes);
 - у примеров есть правильные ответы переменная (response), которую мы предсказываем; она может быть категориальная (categorical), непрерывная или ординальная (ordinal);

- Обучение с учителем (supervised learning) обучение, в котором есть некоторое число примеров с правильными ответами:
 - модель обучается на этой выборке (training phase, learning phase), затем может быть применена к новым примерам (test set);
 - главное обучить модель, которая не только точки из обучающей выборки объясняет, но и на новые примеры хорошо обобщается (generalizes);
 - · иначе оверфиттинг (overfitting);

- Обучение с учителем (supervised learning) обучение, в котором есть некоторое число примеров с правильными ответами:
 - обычно нам дают просто обучающую выборку как тогда проверить, обобщаются ли модели?
 - кросс-валидация разбиваем выборку на тренировочный и валидационный набор (validation set);
 - перед тем как подавать что-то на вход, обычно делают предобработку, стараясь выделить из входных данных самые содержательные аспекты (feature extraction).

- Обучение с учителем (supervised learning) обучение, в котором есть некоторое число примеров с правильными ответами:
 - классификация: есть некоторый дискретный набор категорий (классов), и надо новые примеры определить в какой-нибудь класс;
 - классификация текстов по темам, спам-фильтр;
 - распознавание лиц/объектов/текста;

- Обучение с учителем (supervised learning) обучение, в котором есть некоторое число примеров с правильными ответами:
 - регрессия: есть некоторая неизвестная функция, и надо предсказать её значения на новых примерах:
 - инженерные приложения (предсказать температуру, положение робота, whatever);
 - финансы предсказать цену акций;
 - то же плюс изменения во времени например, распознавание речи.

- Обучение без учителя (unsupervised learning) обучение, в котором нет правильных ответов, только данные:
 - кластеризация (clustering): надо разбить данные на заранее неизвестные классы по некоторой мере похожести:
 - выделить семейства генов из последовательностей нуклеотидов;
 - кластеризовать пользователей и персонализовать под них приложение;
 - кластеризовать масс-спектрометрическое изображение на части с разным составом;

- Обучение без учителя (unsupervised learning) обучение, в котором нет правильных ответов, только данные:
 - снижение размерности (dimensionality reduction): данные имеют огромную размерность (очень много признаков), нужно уменьшить её, выделить самые информативные признаки, чтобы все вышеописанные алгоритмы смогли работать;
 - дополнение матриц (matrix completion): есть разреженная матрица, надо предсказать, что на недостающих позициях.
 - Часто даны правильные ответы для небольшой части данных semi-supervised learning.

- Обучение с подкреплением (reinforcement learning) обучение, в котором агент учится из собственных проб и ошибок:
 - многорукие бандиты: есть некоторый набор действий, каждое из которых ведёт к случайным результатам; нужно получить как можно больший доход;
 - exploration vs. exploitation: как и когда от исследования нового переходить к использованию того, что уже изучил;
 - credit assignment: конфетку дают в самом конце (выиграл партию), и надо как-то распределить эту конфетку по всем ходам, которые привели к победе.

- активное обучение (active learning) как выбрать следующий (относительно дорогой) тест;
- обучение ранжированию (learning to rank) ординальная регрессия, как породить упорядоченный список (интернет-поиск);
- *бустинг* (boosting) как скомбинировать несколько слабых классификаторов так, чтобы получился хороший;
- выбор модели (model selection) где провести черту между моделями с многими параметрами и с немногими.

Вероятность в машинном обучении

- Во всех методах и подходах очень пригодится метод, который мог бы не просто выдавать ответ, а ещё оценивать, насколько модель уверена в этом ответе, насколько модель хорошо описывает данные, как изменятся эти величины при дальнейших экспериментах и т.д.
- Поэтому центральную роль в машинном обучении играет теория вероятностей – и мы тоже будем её активно применять.

Источники

- Christopher M. Bishop, Pattern Recognition and Machine Learning, Springer, 2007.
- Kevin Murphy, Machine Learning: A Probabilistic Perspective, MIT Press, 2013.
- Trevor Hastie, Robert Tibshirani, and Jerome Friedman, *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*, 2nd ed., Springer, 2009.

Байесовский подход

Основные определения

- Нам не понадобятся математические определения сигма-алгебры, вероятностной меры, борелевских множеств и т.п.
- Достаточно понимать, что бывают дискретные случайные величины (неотрицательные вероятности исходов в сумме дают единицу) и непрерывные случайные величины (интеграл неотрицательной функции плотности равен единице).

Основные определения

• Совместная вероятность – вероятность одновременного наступления двух событий, p(x,y); маргинализация:

$$p(x) = \sum_y p(x,y).$$

• Условная вероятность – вероятность наступления одного события, если известно, что произошло другое, $p(x \mid y)$:

$$p(x,y) = p(x \mid y)p(y) = p(y \mid x)p(x).$$

• Теорема Байеса – из предыдущей формулы:

$$p(y|x) = \frac{p(x|y)p(y)}{p(x)} = \frac{p(x|y)p(y)}{\sum_{y'} p(x|y')p(y')}.$$

 \cdot Независимость: x и y независимы, если

$$p(x, y) = p(x)p(y).$$

О болезнях и вероятностях

- Приведём классический пример из классической области применения статистики медицины.
- Пусть некий тест на какую-нибудь болезнь имеет вероятность успеха 95% (т.е. 5% вероятность как позитивной, так и негативной ошибки).
- Всего болезнь имеется у 1% респондентов (отложим на время то, что они разного возраста и профессий).
- Пусть некий человек получил позитивный результат теста (тест говорит, что он болен). С какой вероятностью он действительно болен?

О БОЛЕЗНЯХ И ВЕРОЯТНОСТЯХ

- Приведём классический пример из классической области применения статистики медицины.
- Пусть некий тест на какую-нибудь болезнь имеет вероятность успеха 95% (т.е. 5% вероятность как позитивной, так и негативной ошибки).
- Всего болезнь имеется у 1% респондентов (отложим на время то, что они разного возраста и профессий).
- Пусть некий человек получил позитивный результат теста (тест говорит, что он болен). С какой вероятностью он действительно болен?
- Ответ: 16%.

ДОКАЗАТЕЛЬСТВО

- Обозначим через t результат теста, через d наличие болезни.
- $\cdot \ p(t=1) = p(t=1|d=1)p(d=1) + p(t=1|d=0)p(d=0).$
- Используем теорему Байеса:

$$\begin{split} p(d=1|t=1) &= \\ &= \frac{p(t=1|d=1)p(d=1)}{p(t=1|d=1)p(d=1) + p(t=1|d=0)p(d=0)} = \\ &= \frac{0.95 \times 0.01}{0.95 \times 0.01 + 0.05 \times 0.99} = 0.16. \end{split}$$

Вывод

- Вот такие задачи составляют суть вероятностного вывода (probabilistic inference).
- Поскольку они обычно основаны на теореме Байеса, вывод часто называют байесовским (Bayesian inference).
- Но не только поэтому.

Спасибо!

Спасибо за внимание!