

Как далеко ушли методы анализа данных для объяснения реальности

Малахов Дмитрий

Два типа количественных моделей

- Модели для построения прогнозов

- Модели для выявления причинно-следственной взаимосвязи

Что такое вообще это машинное обучение?

- Роботы захватят мир?
- Нейросети – это система уравнений, которые передают друг другу сигналы и вместе формируют отклик системы (число)
 - Такие модели почему-то очень быстро и аккуратно выявляют закономерности в данных и их потом экстраполируют
- Уравнения цели не имеют, но ракета, управляемая уравнением, цель имеет
 - У искусственного “интеллекта” нет осознанности...
- Машинное обучение состоит не только из нейросетей...

Что привело к популярности моделей машинного обучения

- Рост объема данных
 - Цифровизация экономики
 - Сохранение и обработка большого количества данных
- Увеличение компьютерных мощностей
 - Увеличение скорости вычислений
 - Увеличение размера памяти компьютеров для сохранения больших объемов данных
- Значительные успехи в поиске максимума сложных функций

Достоинства и недостатки машинного обучения

- Машинное обучение позволяет выявлять нелинейные и сложные взаимосвязи
- Где машинное обучение особенно успешно?
 - Машинный перевод и создание текстов
 - Распознавание (и создание) изображений
 - Распознавание (и создание) звуков
 - Распознавание (и создание) видео
- Это области, где теоретических знаний мало, но можно оцифровать входные данные.
 - Например, картинка – блок (матрица), состоящий из цифр, отвечающих за цвет пикселя
- НО (!) модели машинного обучения слабо поддаются интерпретации, это черные ящики. Многие ученые не считают машинное обучение наукой в строгом ее понимании

Каузальные модели

- Модели для тестирования причинно-следственных взаимосвязей гораздо более простые. Но они интерпретируемы!
- Идеалом в данном случае является двойной “слепой” рандомизированный плацебо контролируемый эксперимент
- Исходя из его результатов, можно сказать влияет ли X на Y или нет
- Могут ли бы быть эксперименты в социальных науках?

Текущий тренд в количественном анализе

- Совмещение методики анализа экспериментов и “black box” моделей
 - Human decisions and machine predictions (2017) – может ли компьютер (модель машинного обучения) быть более объективным и точным, чем человек судья?
- Сейчас мы можем оцифровать и квантифицировать почти любое человеческое действие
- При этом есть проблемы, например, с формированием полной и подробной базы НКО в РФ. Может ли здесь помочь машинное обучение?
 - Да, например, при расшифровке отчетов, которые сохраняются как изображения