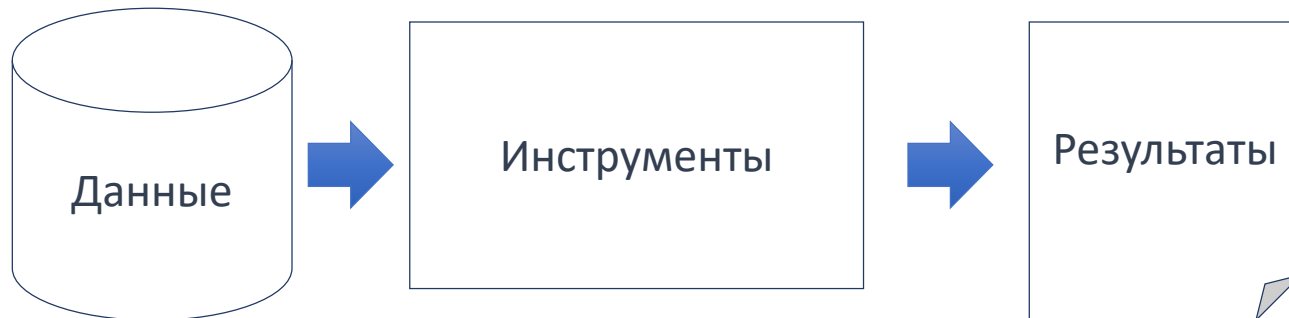


cos & ht

Применение искусственного интеллекта для работы с документами. Чего ждать и чего не ждать на этом минном поле.

Общая постановка задачи



Классификация типов данных

Структурированные данные

- XML документы
- записи в информационных системах
- информация получаемая по запросам

Неструктурированные данные

- Сканированные документы
- Документы в офисных форматах
- Информация из сети интернет и социальных сетей



**ФЕДЕРАЛЬНОЕ АГЕНТСТВО
ПО ПЕЧАТИ И МАССОВЫМ
КОММУНИКАЦИЯМ
(РОСПЕЧАТЬ)**

ЗАМЕСТИТЕЛЬ РУКОВОДИТЕЛЯ

Страстной бульвар, д.5, Москва ГСП-4, 127994
Справочная: (495)694-11-77. Факс: (495)694-22-81
E-mail: info@fapmc.ru
<http://www.fapmc.ru> <http://rospechat.ru>
11.10.2019 № 00000016

На № от

Об обмене документами в электронном виде с применением электронной подписи

Федеральное агентство по печати и массовым коммуникациям в соответствии с письмом Министерства цифрового развития, связи и массовых коммуникаций Российской Федерации от 17.09.2016 г. № ОП-П19-070-19529 «Об обмене документами в электронном виде с применением электронной подписи» направляет отчет Роспечати об обмене документами в электронном виде, подписанными электронной подписью, за сентябрь 2019 г.

Приложение: Статистический отчет Федерального агентства по печати и массовым коммуникациям об обмене документами в электронном виде, подписанными электронной подписью, за сентябрь 2019 г. на 1 л. в 1 экз.

Т.В.Наумова

Исполнитель: Каширин А. А., тел. +7 (495) 6941 547

<?xml version="1.0" encoding="UTF-8" ?>

<Document>

<Addressy medo_id="001">

Министерство цифрового развития, связи и массовых коммуникаций Российской Федерации (Минкомсвязь России)

</Addressy>

<Sourcemedo_id="002"> ФЕДЕРАЛЬНОЕ АГЕНТСТВО
ПО ПЕЧАТИ И МАССОВЫМ
КОММУНИКАЦИЯМ
(РОСПЕЧАТЬ)

</Source>

<author id="12">ЗАМЕСТИТЕЛЬ РУКОВОДИТЕЛЯ</author>

<address>Страстной бульвар, д.5, Москва ГСП-4, 127994</address>

<phone>+7(495)694-11-77</phone> <mail>info@fapmc.ru</mail>

<url><http://www.fapmc.ru></url>

<date>11.10.2019</date>

<num> 00000016</num>

<text>

Об обмене документами в электронном виде с применением подписи Федеральное агентство по печати и массовым коммуникациям в соответствии с письмом Министерства направляет отчет Роспечати об обмене документами в электронном виде, подписанными электронной подписью, за сентябрь 2019 г.

</text>

<attach>Статистический отчет Федерального агентства по печати и массовым

коммуникациям об обмене документами в электронном виде, подписанными

электронной подписью, за сентябрь 2019 г. на 1 л. в 1 экз.

</attach>

<sign>Т.В.Наумова</sign>

<resp>Каширин А. А.</resp>

<phone> +7 (495) 6941 547</phone>

<cert>

<SN>7C25B9409F55379F81161AD023AB72FE220124DD</SN>

<CN>Наумова Татьяна Владимировна</CN>

</cert>

</Document>

Виды машинного обучения

Обучение с учителем
(supervised learning)

Требует большого объема работы большого числа экспертов

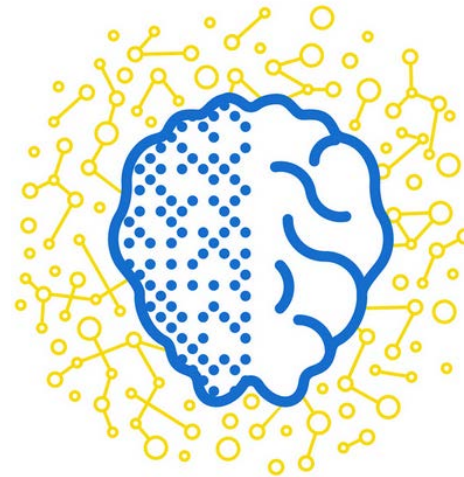
Обучение без учителя или самообучение
(unsupervised learning)

Требует хорошей инфраструктуры и наличия истории. Позволяет находить неочевидные для экспертов связи. Тоже, по – видимому, хорошо реализуемо.

Обучение с подкреплением
(reinforcement learning)

А вот это появится в тот момент, когда будет решена задача автоматического отлова нарушений и даст возможность упреждающего анализа информации

Примеры применения нейросетей



Разбор сообщений граждан



Санкт-Петербургский
информационно-аналитический центр

Демонстрационный стенд AIDOC.ЖКХ

Стенд позволяет определить тематику обращения граждан по проблемам ЖКХ. Введите сообщение в поле "ввод текста" и получите результат классификации в течение 30 секунд.

Ввод текста

На углу Невского и Литейного большая яма с водой. Проезжающие машины обливают пешеходов

Отправить

Результат классификации

Номер категории: 9
Категория: Улица / Благоустройство / Повреждения дорожного покрытия проезжей части

Определение тематики сообщений - на основе классификатора ЖКХ г. Санкт-Петербурга.



Задача автоматической классификации документов

Входящие письма некоторого агентства обрабатываются отделом документального обеспечения, аннотируются и направляются для дальнейшей обработки руководителю или одному из его заместителей по тематической принадлежности.

Руководитель назначает список сотрудников, с персонализированными резолюциями для каждого.

Нейронная сеть позволяет


- Автоматически аннотировать входящий документ в соответствии с набором заданных семантик
- Автоматически выбрать адресата из числа руководителя и четырех его заместителей
- Автоматически сформировать список исполнителей по входящему документу и расписать резолюции


Федеральное агентство по печати
и массовым коммуникациям




Антиплагиат

- Решается задача поиска документов, из которых имеются прямые заимствования

Более
· · · **300** млн. источников
 Самая полная русскоязычная коллекция источников в России и СНГ

От
· · · **10** сек. на обработку
 Моментальная проверка документа

Ровно
· · · **3** шага
 до получения результата

От
· · · **0** руб.
 Возможность бесплатной проверки

Примеры применения аналитики. Семантический разбор договоров



Решается задача выделения структур внутри неструктурированных текстов

- **Поиск реквизитов и адресов сторон**
- **Сроки выполнения работ**
- **Сроки оплат**
- **Штрафные санкции**

Типы решаемых задач

В данном разделе перечислены основные решаемые задачи и проведена их привязка к используемым инструментам (как традиционным, так и новым)

Задачи классификации

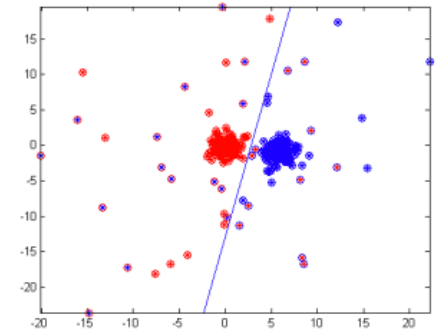
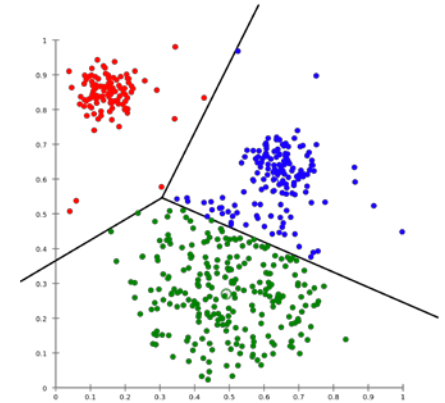
Задача классификации — задача, в которой имеется множество объектов (ситуаций), разделённых некоторым образом на классы.

Необходимо для объекта определить, к какому классу он относится.

Типичная классификация — «хороший объект» vrs «плохой объект»

Например:

- «Контракт исполнен по правилам»
- «Контракт исполнен с нарушениями»



Машинное обучение и AI

Машинное обучение — это «[...] класс методов искусственного интеллекта, которые позволяют улучшить результаты работы компьютеров путем обучения на известных данных», — [Berkeley](#).



Фото с котом



Фото без кота



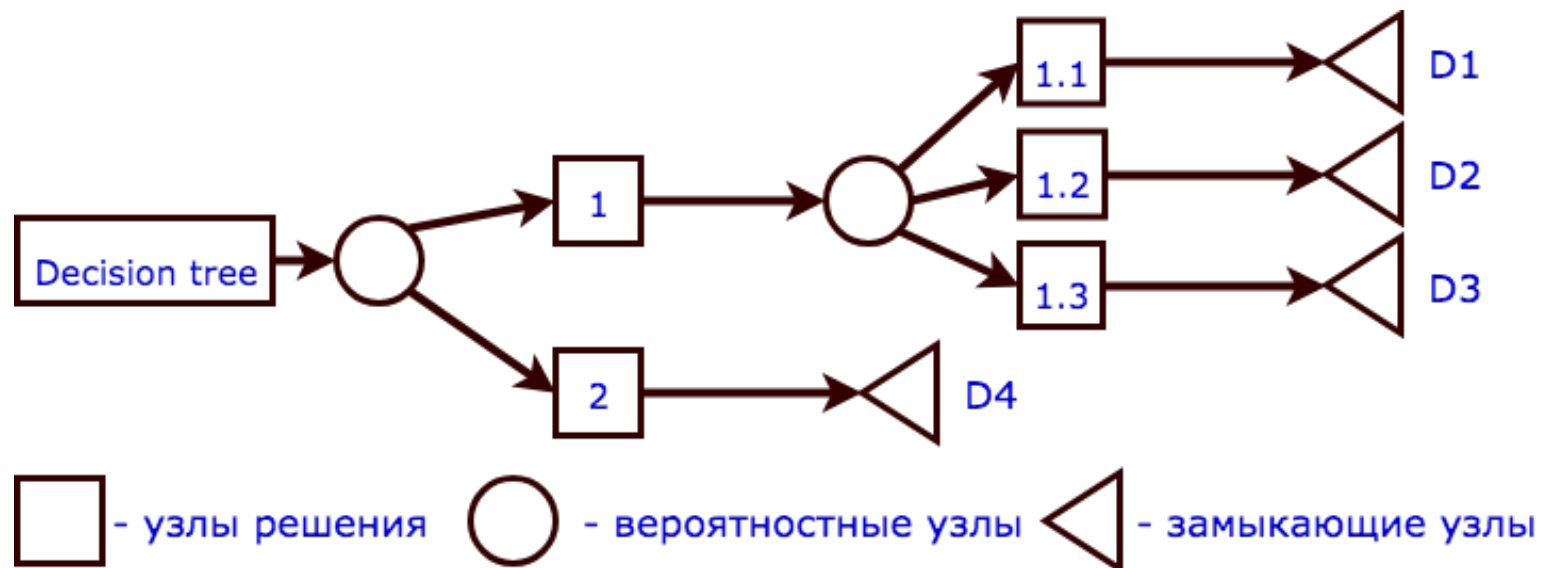
Контракт без признаков коррупции



Контракт с признаками коррупции

Задачи поиска аналогов в прошлом

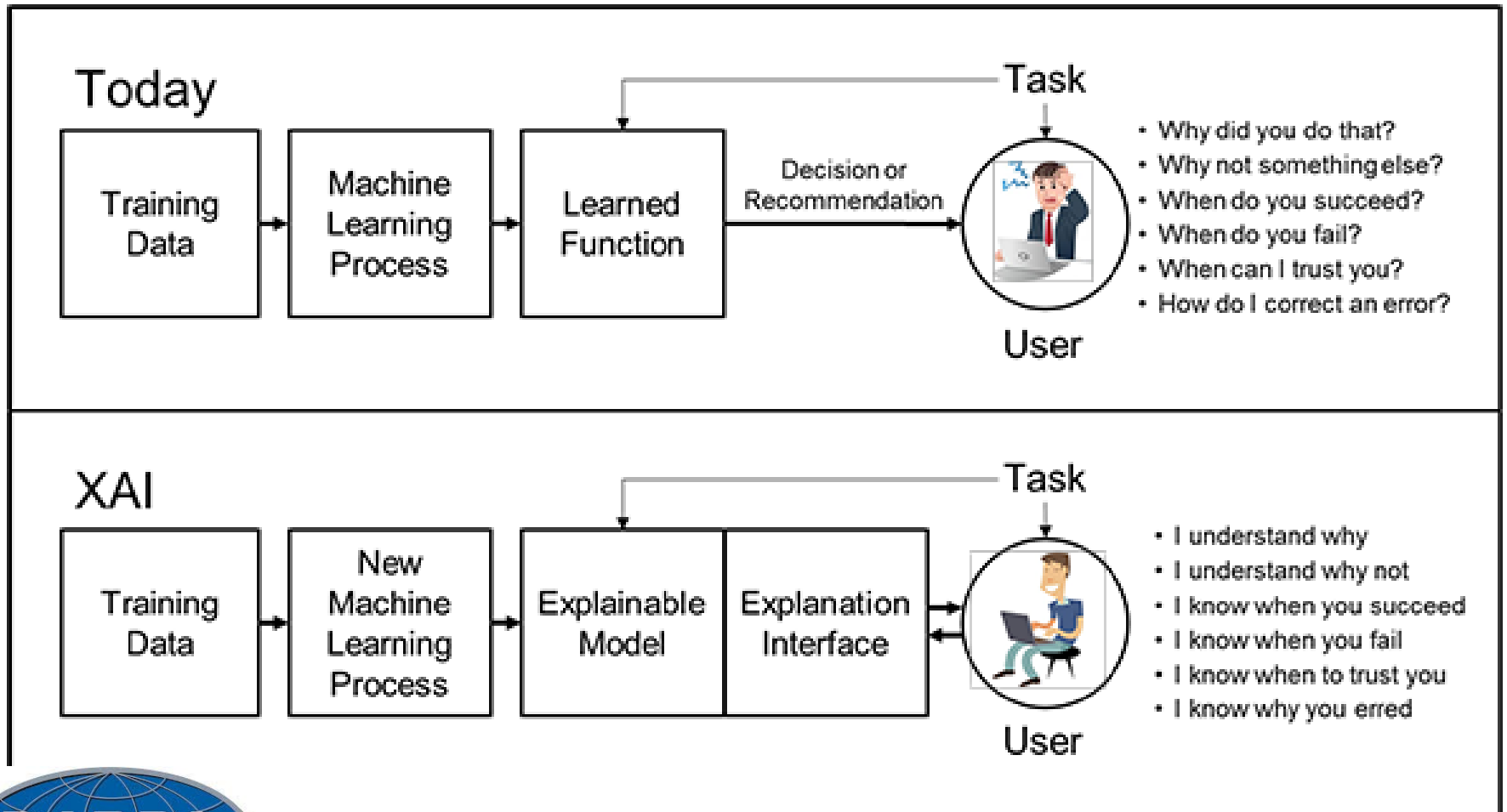
Самый известный способ –
Дерево принятия решений



Обещанные минные поля

- Нейросеть принимает решение но объяснить на основании чего она не может
- Ее легко можно научить плохому
- Иногда создается ощущение что ИИ придумали производители HW
- А не разучимся ли мы управлять?
- И даже правам человека нашлось место

Нейросеть принимает решение но объяснить на основании чего она не может



Ее легко можно научить плохому

- В городе N при распределении контрактов на застройку всегда побеждала компания «Лютик» хорошего знакомого городского главы
- На архиве решений за 3 года обучили нейросеть
- На очередных выборах поменяли городского главу
- А нейросеть продолжила выдавать контракты компании «Лютик»

... или забыть научить хорошему

- Назначили нового начальника управления Сидорова
- Но ИИ не дурак – если нет Иванова – его заменяет Петров
- А Сидоров никак не поймет , почему ИИ не дает ему работать

Иногда создается ощущение что ИИ придумали производители HW.



776 510

- Обработка 1 000 000 записей по ЖКХ на NVIDIA TITAN RSX заняло практически месяц
- NVIDIA® TITAN RTX™ – самая быстрая из когда-либо созданных видеокарт для ПК. Она работает на базе удостоенной наград архитектуре Turing™, обеспечивает производительность 130 Терафлопс и заряжает ваш ПК мощностью 576 тензорных ядер и 24 Гб ультраскоростной памяти GDDR6.

А не разучимся ли мы управлять?



«Никакого опыта у летчиков, летающих сегодня на Boeing и Airbus, где есть программа, практически нет. Летчики разучились ручному визуальному пилотированию»

И даже правам человека нашлось место



По словам Федотова, все, что будет изучать кафедра, связано с правами человека. Он привел в пример ситуацию, когда клиент авиакомпании звонит по вопросам покупки билетов, а ему отвечает робот, не способный дать ответы на некоторые вопросы. «Человек человека всегда поймет, а искусственный интеллект — нет», — пояснил глава СПЧ. Он добавил, что кафедра будет уделять большое внимание искусственному интеллекту: «Для правоведа это новая галактика».