



Диалоговые системы: State-Of-The-Art и реальность

Ирина Пионтковская,
Московский Физико-Технический Институт
Проект iPavlov <https://ipavlov.ai>



CONVERSATIONAL
ARTIFICIAL
INTELLIGENCE

Почему это СТРАШНО



Кошмар product manager'a

Ругается матом
Говорит глупости
Выдает персональные данные
=> проблемы вплоть до
судебных исков

Почему это СТРАШНО



Кошмар project manager'a

Процесс выходит из-под контроля
Ученые делают не то что ожидается
Результат не такой как хочется

=> неудачный проект

Почему это СТРАШНО



Кошмар IT отдела

Работает медленно
Пожирает вычислительные ресурсы
Не отлаживается

=> проблемы в поддержке

Почему это СТРАШНО



Кошмар Сотрудников

Заменяет людей:

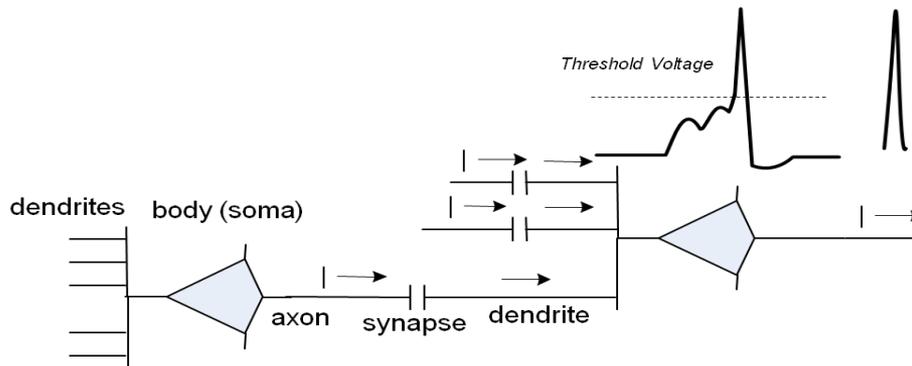
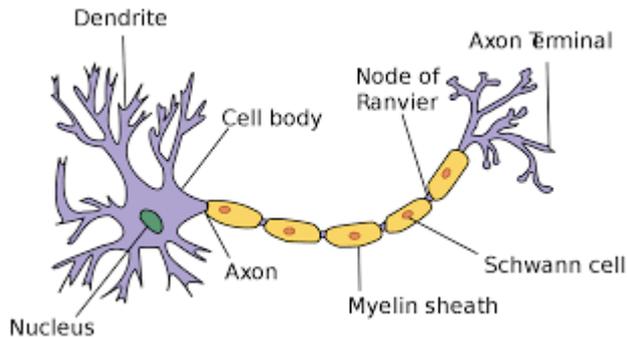
- в колл-центрах
- службе поддержки...

=> нас всех уволят!!!!

Что такое AI



Математическая модель мозга, которая давно живет своей жизнью



Что такое AI



Алгоритмы, создающие другие алгоритмы на основе данных

Training



Inference

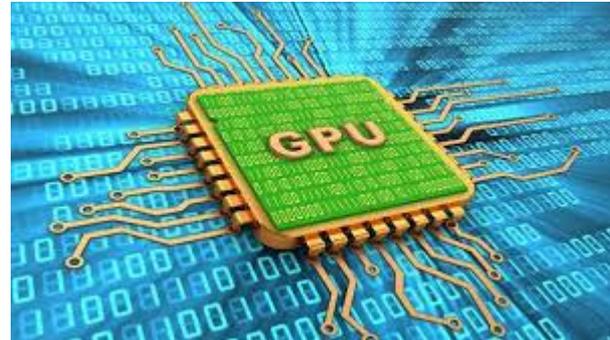
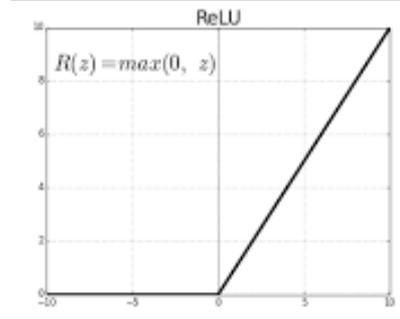
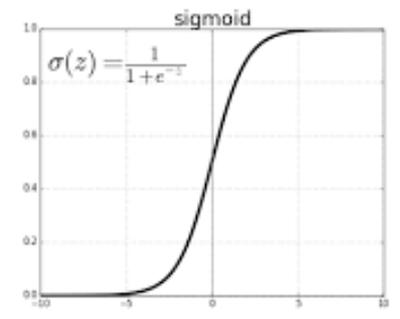


dog

Что такое AI



Умножение матриц, применение непрерывных функций и стохастический градиентный спуск



Что такое AI



Свойства:

- Элемент случайности при обучении –
не всегда предсказуемый результат
- Миллиарды параметров –
высокая вычислительная сложность
- Для обучения требуется большое количество данных, в большинстве случаев с ответами –
трудно создавать датасеты

Что умеем и чего не умеем



Новый класс алгоритмов, который во многих случаях гораздо лучше справляется с «нелогичными» задачами

Генерация текста

* не держит общую мысль текста

daybreak
a cold snow falls
into the sound of the bay

Что умеем и чего не умеем



Новый класс алгоритмов, который во многих случаях гораздо лучше справляется с «нелогичными» задачами

Ответы на вопросы

* Находит трудные неочевидные ответы, но упускает простую логику

Mary went to the bathroom.
John moved to the hallway.
Mary travelled to the office.
Where is Mary? A: office

Что умеем и чего не умеем



Новый класс алгоритмов, который во многих случаях гораздо лучше справляется с «нелогичными» задачами

Семантический поиск

* Может перепутать синонимы и антонимы

HUGE synonyms:	large	0.337
	massive	0.395
	vast	0.446
	small	0.531

Что умеем и чего не умеем



Новый класс алгоритмов, который во многих случаях гораздо лучше справляется с «нелогичными» задачами

Ведение диалога

*Нет стабильной «личности», может давать противоречивую информацию

Сколько тебе лет?

Я в 7 классе

Чем ты увлекаешься?

У меня трое детей

Правила vs. Deep Learning

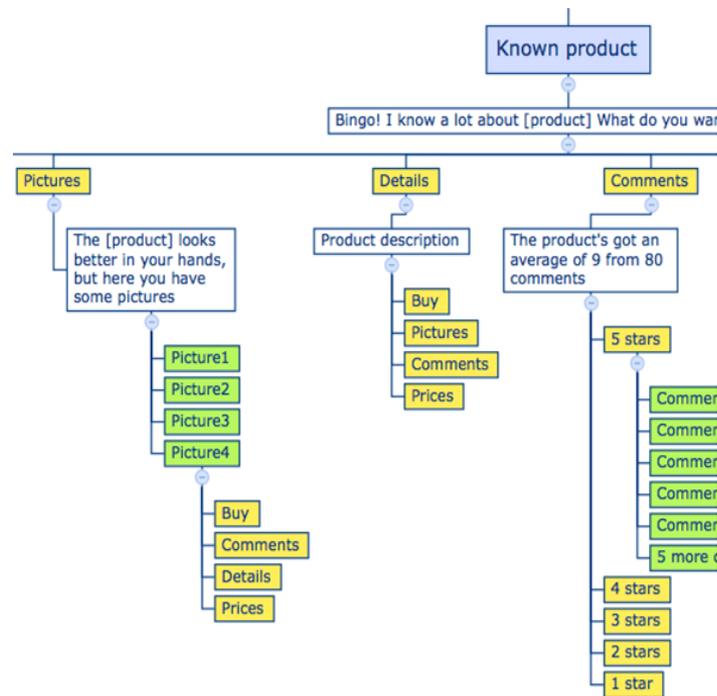


Правила:

- ❑ Логичны и Предсказуемы
- ❑ Легко исправить Одну Конкретную Ошибку
- ❑ Хорошо справляются только с простыми случаями

Deep Learning:

- ❑ В среднем, гораздо лучше справляются с пониманием текста
- ❑ Однажды обученную модель трудно изменить; практически невозможно изменить поведение на одном конкретном примере
- ❑ Более похожи на человеческий разговор



Кому нужен AI



Естественный отбор

Половой отбор



Приспособление к среде обитания



Кому нужен AI



- Нет готовых решений “из коробки”, но есть классные составляющие
- Высокий риск неудачи, но ожидаемый выигрыш значителен:
 - в некоторых случаях AI алгоритмы дают существенное улучшение качества системы
 - в некоторых случаях «wow-эффект»

Существует огромный запрос на AI решения (где угодно)

Usecase: Question Answering system for Oil & Gas Documentation



iPavlov



DATA
MONSTERS

Постановка задачи



Вопросно-ответная система по документации для компании-поставщика нефтяного оборудования

- ❖ Очень специфическая лексика
- ❖ Много данных в виде таблиц

Control the pH at 10.5 or above using lime. Do not use caustic soda alone, because caustic reacts to form sodium sulfide (Na_2S) which is extremely soluble and H_2S may be released from the fluid if the pH drops.

Always start the pumps before the rotary. In narrow equivalent circulating density (ECD) margin wells it may be necessary to start rotary before bringing the pumps up. In such situation, ensure that the pumps are started as soon as possible.

Понимание текста



Ответы на вопросы по тексту -- одна из задач Reading Comprehension (понимания текста)

- Контекст:

In meteorology, precipitation is any product of the condensation of atmospheric water vapor that falls under gravity. The main forms of precipitation include drizzle, rain, sleet, snow, graupel and hail... Precipitation forms as smaller droplets coalesce via collision with other rain drops or ice crystals **within a cloud**. Short, intense periods of rain in scattered locations are called “showers”.

- Вопрос:

Where do water droplets collide with ice crystals to form precipitation?

Датасет SQuAD



- Stanford Question Answering Dataset ([SQuAD](#))
 - контексты собраны из википедии
 - ~ 23k контекстов из 536 случайных статей из топ 10k
 - ~ 100k вопросов
 - английский язык

Метрики:

EM - exact match, точное совпадение

F1 - мера пересечения по словам

SQuAD Leaderboard

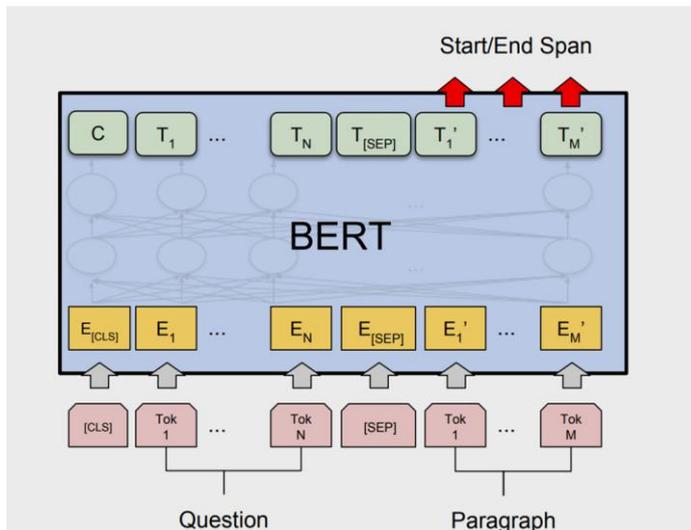


Rank	Model	EM	F1
	Human Performance <i>Stanford University</i> (Rajpurkar et al. '16)	82.304	91.221
1 Oct 05, 2018	BERT (ensemble) <i>Google AI Language</i> https://arxiv.org/abs/1810.04805	87.433	93.160
<hr/>			
66 Aug 27, 2016	Match-LSTM with Ans-Ptr (Sentence) <i>Singapore Management University</i> https://arxiv.org/abs/1608.07905	54.505	67.748

BERT



BERT -- предобученная модель от Google для векторных представлений текстов (предложений)



Задачи для предобучения (на неразмеченных данных):

- Предсказание пропущенного слова в предложении
- Предсказание следующего предложения

64 TPU чипа, обучение заняло 4 дня для каждой задачи

Ответы на вопросы



SQuAD 2.0 -- обновленный вариант датасета, содержащий 50000 вопросов, на которые ответа нет. Теперь система умеет определять, когда ответа нет

	Human Performance <i>Stanford University</i> (Rajpurkar & Jia et al. '18)	86.831	89.452
1 Sep 13, 2018	nlnet (single model) <i>Microsoft Research Asia</i>	74.238	77.022

Ответы на вопросы



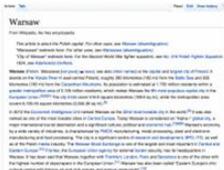
Вопросно-ответная система: поиск нужного абзаца в тексте + поиск ответа в абзаце (Open Domain Question Answering)

Q: How many of Warsaw's inhabitants spoke Polish in 1933?



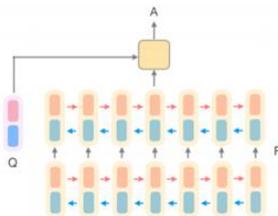
WIKIPEDIA
The Free Encyclopedia

Document
Retriever



Document
Reader

833,500



Best results:

SQuAD-Open

F1 42.5

EM 34.7

Figure 1: An overview of our question answering system DrQA.

Ответы на вопросы по таблицам



WikiSQL -- датаст, подготовленный Salesforce
~80 000 вопросов
~25 000 таблиц из википедии
Задача: перевести вопрос в SQL запрос

Best result:
87.1%
execution accuracy

Table: CFLDraft

Pick #	CFL Team	Player	Position	College
27	Hamilton Tiger-Cats	Connor Healy	DB	Wilfrid Laurier
28	Calgary Stampeders	Anthony Forgone	OL	York
29	Ottawa Renegades	L.P. Ladouceur	DT	California
30	Toronto Argonauts	Frank Hoffman	DL	York
...

Question:

How many CFL teams are from York College?

SQL:

```
SELECT COUNT CFL Team FROM  
CFLDraft WHERE College = "York"
```

Result:

2

Выделение сущностей



What is the **drill rig motor** **5GEB22D** **voltage**?

equipment

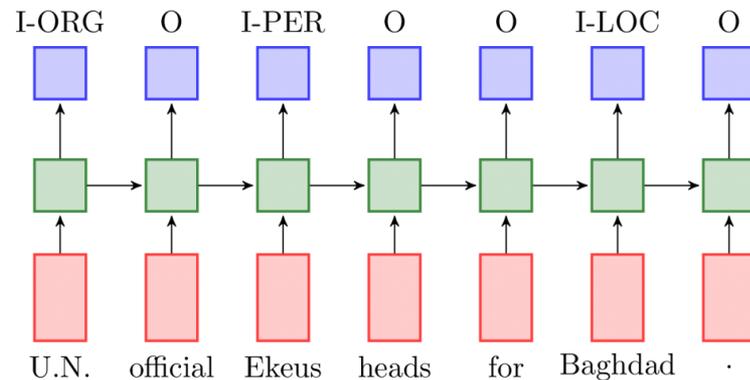
model

property

Стандартная задача:

Person, Location, Organization, Misc

OntoNoted dataset: ~80000
сущностей

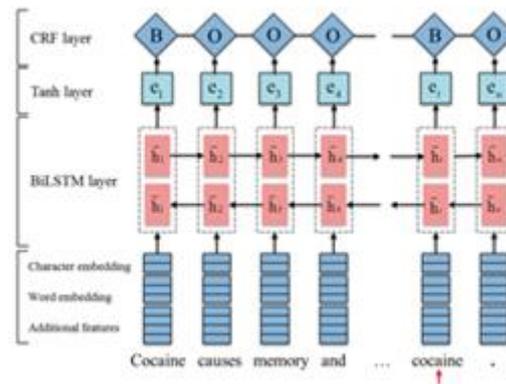


Как уменьшить количество разметки



Автоматическое выделение терминов --
разметка начинается не с нуля; ассессорам
приходится только поправлять выделенные
термины

Использование предобученной языковой
модели позволяет обучать выделение
сущностей на ~100 размеченных текстах



<https://github.com/dialogue-evaluation/factRuEval-2016>

FactRuEval: 88.7 F1

На создание модели ушло ~ 2 недели

THANK YOU

