Международный военно-технический форум АРМИЯ 2022
заседание секции №3 «Научная проблематика в области искусственного интеллекта»

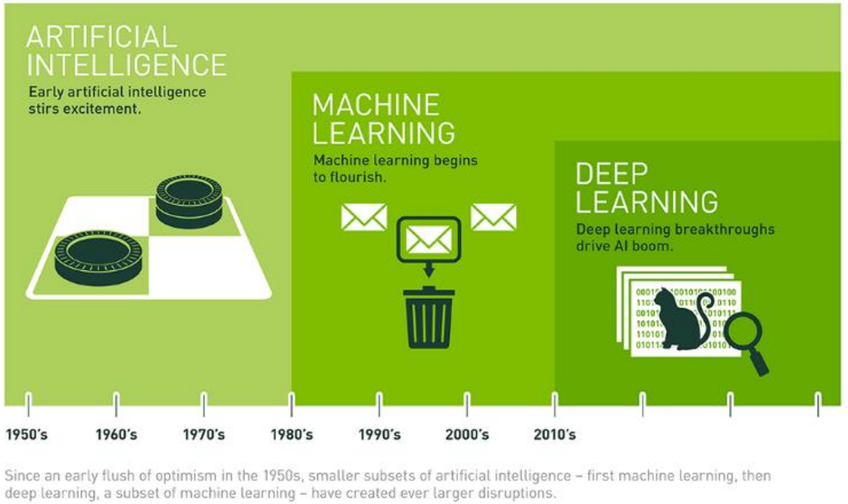
Обучаемая векторизация данных как основа нейросетевых технологий искусственного интеллекта

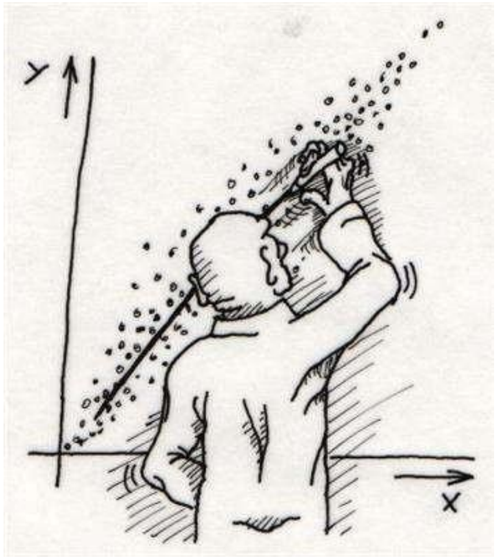
Воронцов Константин Вячеславович

д.ф.-м.н., профессор РАН,
зав. лаб. Машинного обучения и семантического анализа
Института Искусственного Интеллекта МГУ,
профессор, и.о. зав. кафедрой Математических методов прогнозирования ВМК МГУ,
г.н.с. ФИЦ «Информатика и управление» РАН,
профессор, зав. кафедрой МФТИ

Машинное обучение (Machine Learning, ML)

- одна из ключевых информационных технологий будущего
- наиболее успешное направление ИИ, вытеснившее экспертные системы и инженерию знаний





- проведение функции через заданные точки в сложно устроенных пространствах
- математическое моделирование в условиях, когда знаний мало, данных много
- тысячи различных методов и алгоритмов
- более 100 000 научных публикаций в год

Задача машинного обучения с учителем

Этап №1 – обучение (train)

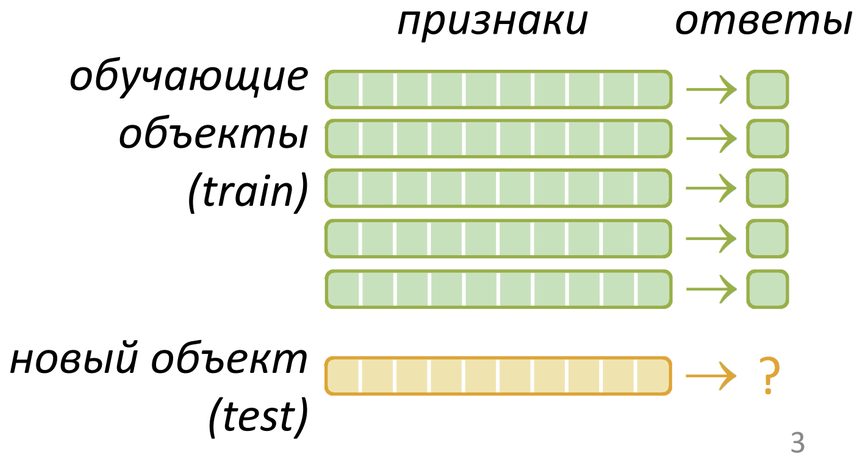
- **На входе:**
  данные – выборка пар «*объект* → *ответ*»,
  каждый объект описывается *вектором признаков*
- **На выходе:**
  модель, предсказывающая ответ по объекту

Задача поставлена,
если у неё есть «ДНК»:

- ***Дано***
- ***Найти***
- ***Критерий***

Этап №2 – применение (test)

- **На входе:**
  данные – новый объект
- **На выходе:**
  предсказание ответа на новом объекте



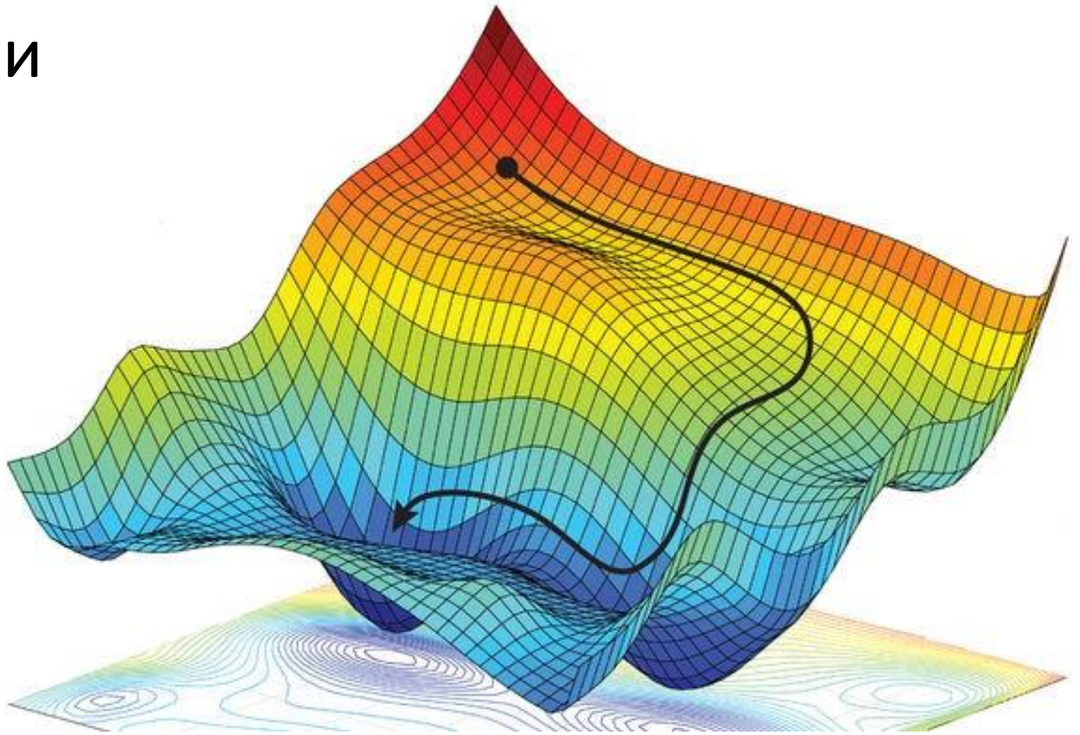
Машинное обучение – это оптимизация

x – вектор объекта обучающей выборки

w – параметры модели

$\mathrm{Loss}(x, w)$ – функция потерь

$Q(w)$ – критерий качества модели



Задача на этапе обучения модели:

$$Q(w) = \sum_{x} \mathrm{Loss}(x, w) \to \min$$

Способ решения – численные методы оптимизации

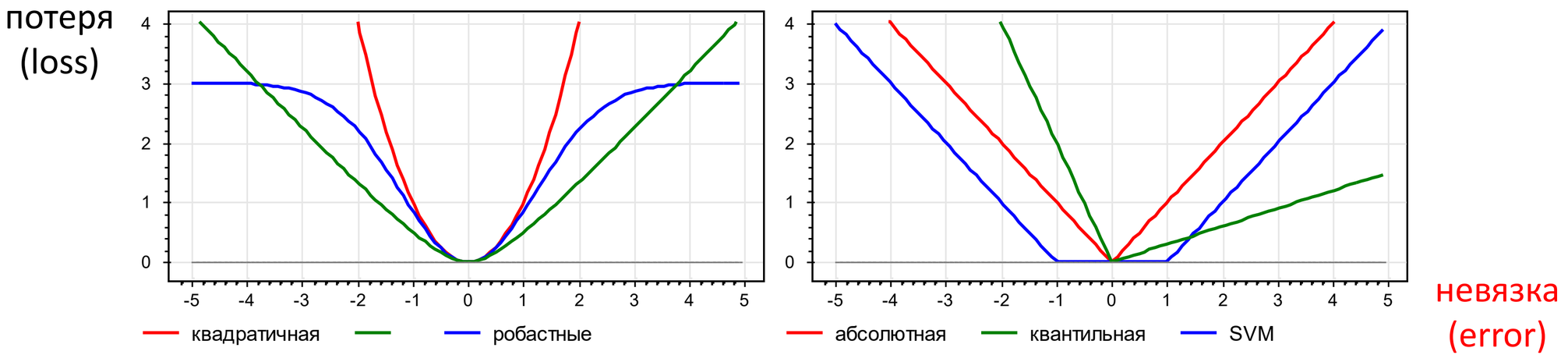
Восстановление регрессии (regression)

x – вектор объекта обучающей выборки, y – числовой ответ

$a(x, w)$ – модель регрессии с параметрами w

Например, $a(x, w) = \sum_j w_j x_j$ — линейная модель регрессии

$\text{Loss}(x, w) = (a(x, w) - y)^2$ – квадратичная функция потерь



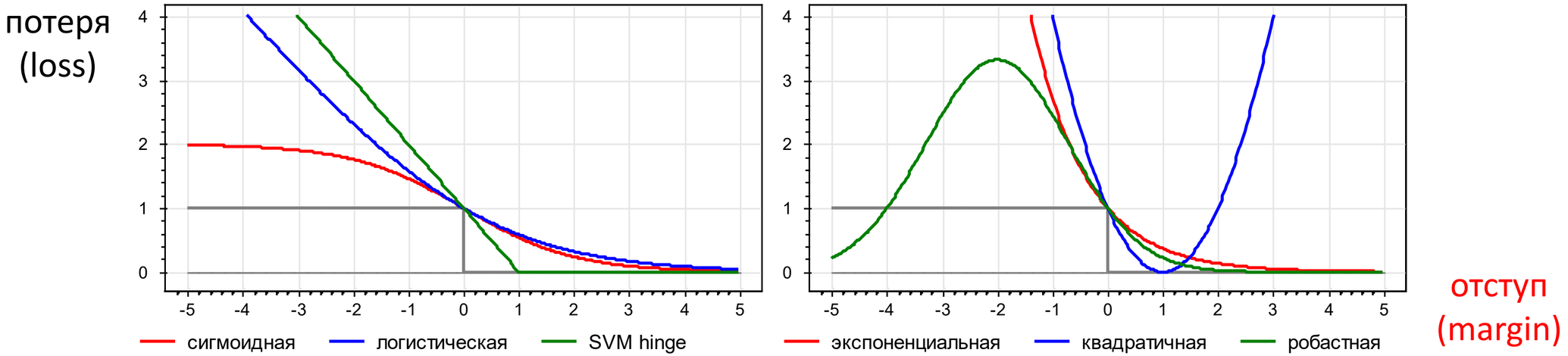
Классификация (classification)

x – вектор объекта обучающей выборки, y – ответ ($+1$ или -1)

$a(x, w)$ – модель классификации с параметрами w

Например, $a(x, w) = \text{sign}\left(\sum_j w_j x_j\right)$ — линейная модель

$\text{Loss}(x, w) = \max\left(0, 1 - y\sum_j w_j x_j\right)$ – функция потерь hinge

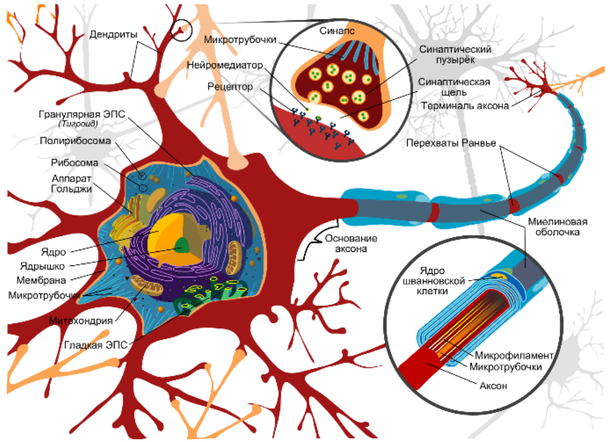


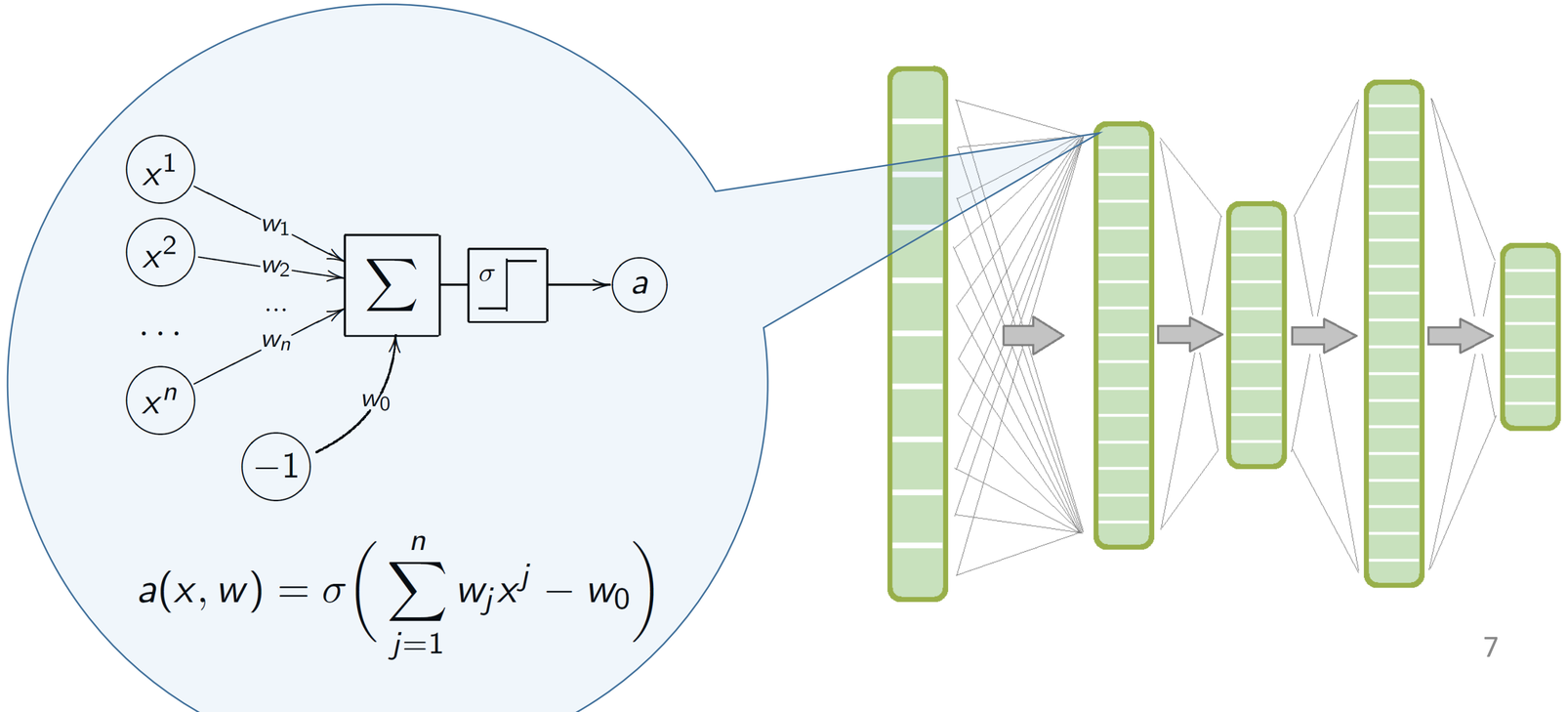
Искусственные нейронные сети

На каждом слое сети вектор объекта преобразуется в новый вектор

Эти преобразования обучаемые, их параметры входят в w

Каждое преобразование (нейрон) – взвешенная сумма признаков

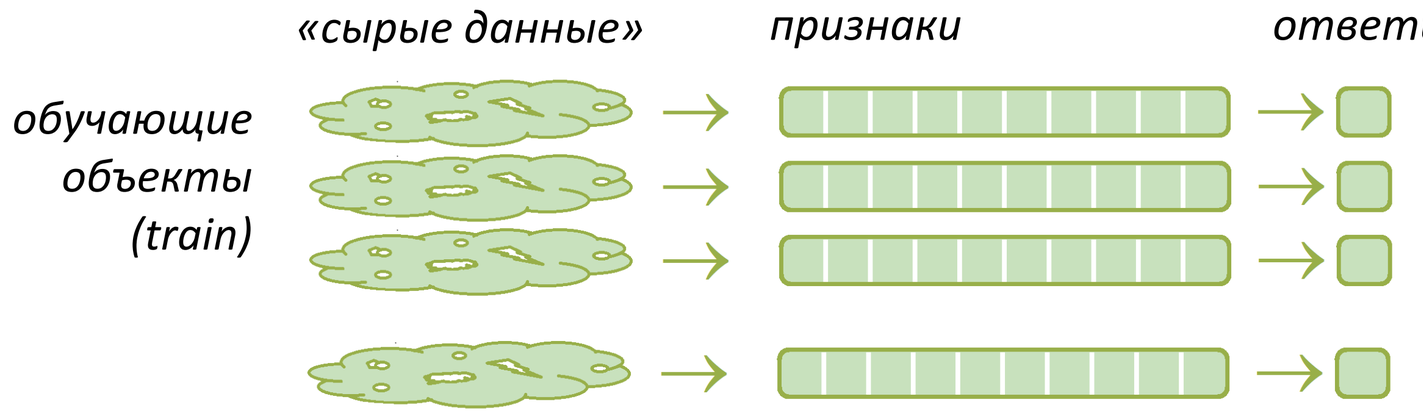




Глубокие нейронные сети

Вход: сложно структурированные «сырые» данные объектов

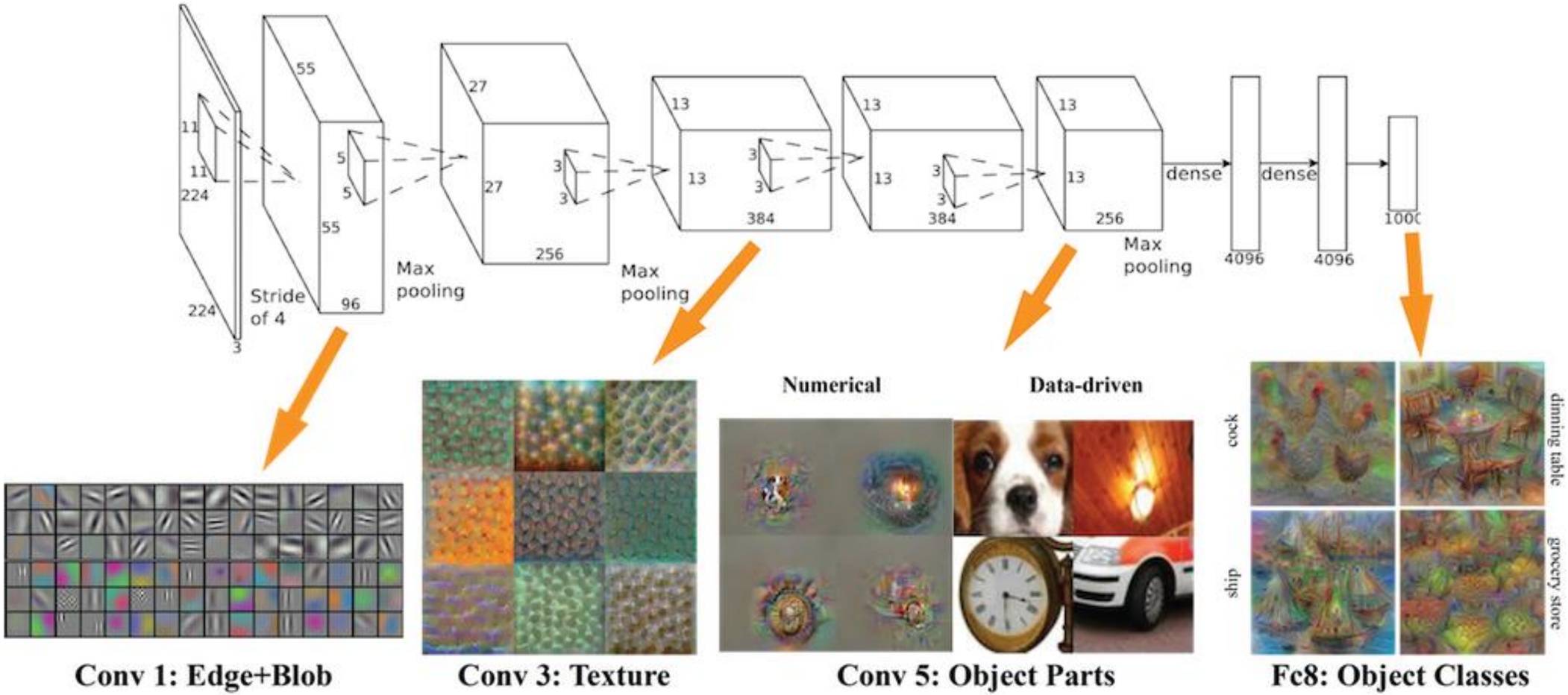
Выход: ответы, векторные представления объектов



Deep Learning – это обучаемая векторизация сложных объектов

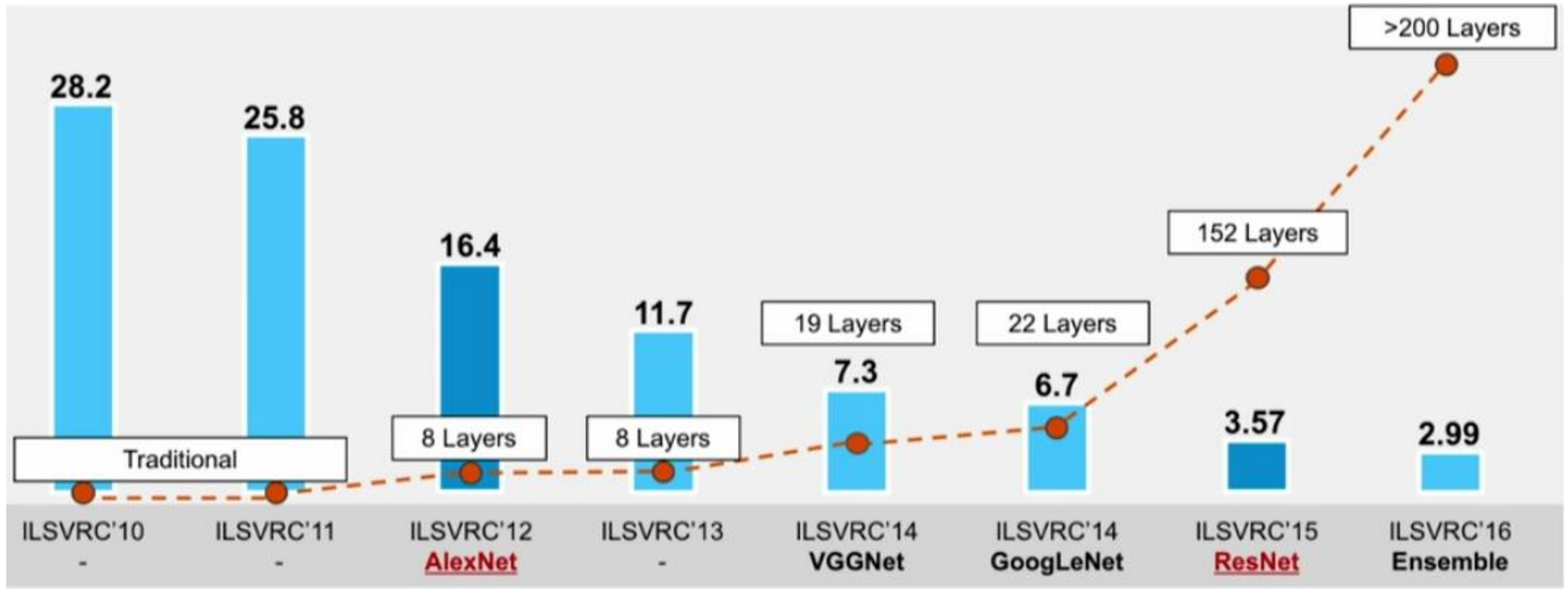
Примеры сложно структурированных объектов:
тексты, изображения, видео, временные ряды, транзакции, графы, ...

Глубокие свёрточные нейронные сети
для классификации изображений



Роль больших данных

ImageNet: открытая выборка 14M изображений, 20K категорий



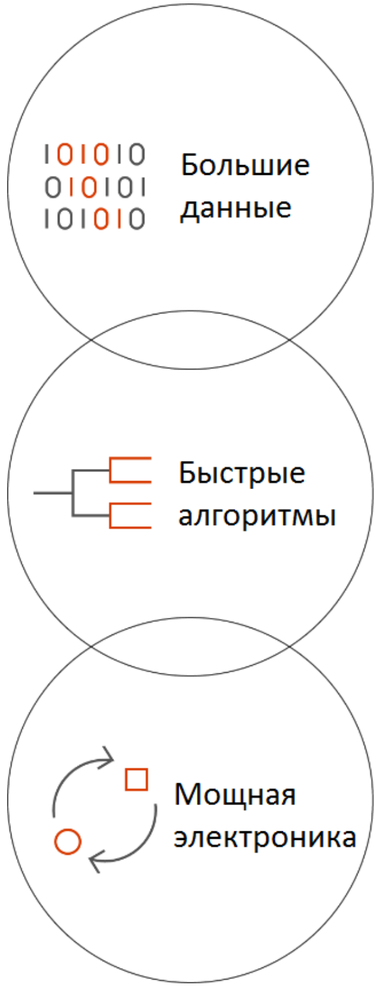
Старт в 2009 г. Человеческий уровень ошибок 5% пройден в 2015 г.

Три составляющих успеха Deep Learning

- Повсеместное применение компьютерных технологий
  → *накопление больших выборок данных*
  в частности, ImageNet

- Развитие математических методов и алгоритмов
  → *накопление критической массы опыта*
  методы оптимизации, контроль переобучения

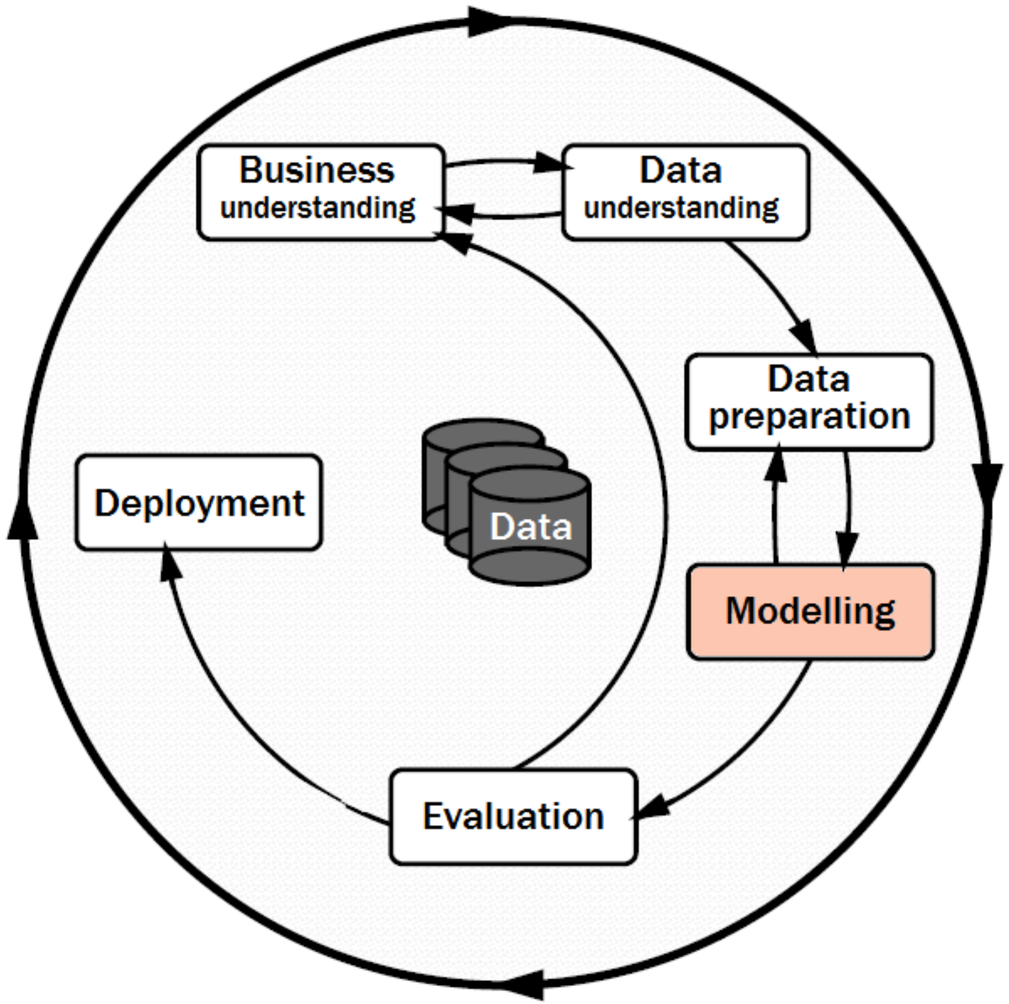
- Достижения микроэлектроники
  → *рост вычислительных мощностей по закону Мура*
  в частности, GPU



Этапы решения задач ML/DS/AI

CRISP-DM: CRoss Industry Standard Process for Data Mining (1999)

(SPSS, Teradata, Daimler AG, NCR Corp., OHRA)



- понимание прикладной задачи
- понимание данных
- конструирование признаков
  → обучаемая векторизация (DL)
- обучаемые модели (ES → ML)
- оценивание решения
  → AutoML
- внедрение и эксплуатация
  → бесшовная эволюция моделей (RL)

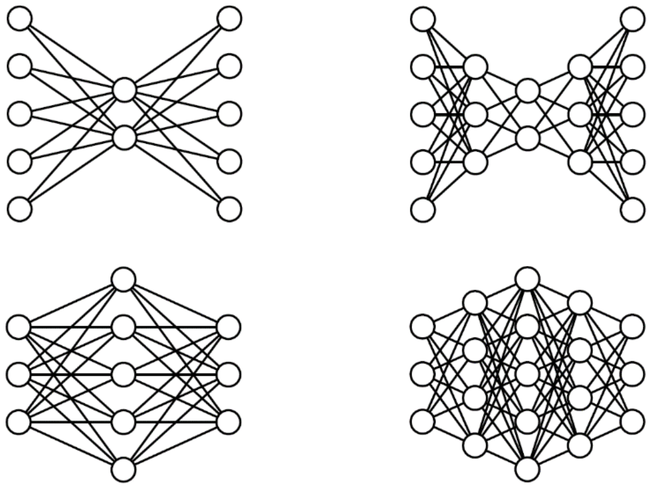
Обучаемая векторизация (autoencoders)

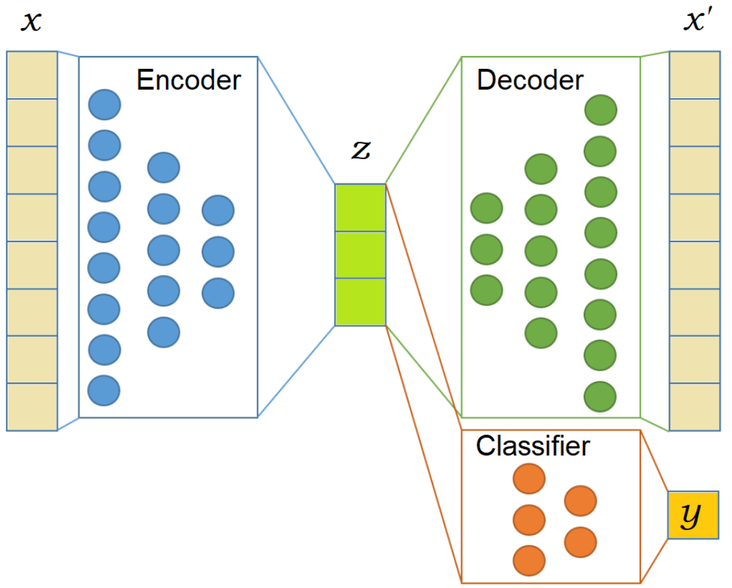
x – вектор объекта обучающей выборки, ответов не дано

$z = f(x, w)$ – модель кодирования x в векторное представление z

$x' = g(z, w)$ – модель декодирования z в реконструкцию x'

$\text{Loss}(x, w) = \|x'(w) - x\|$ – точность реконструкции объекта





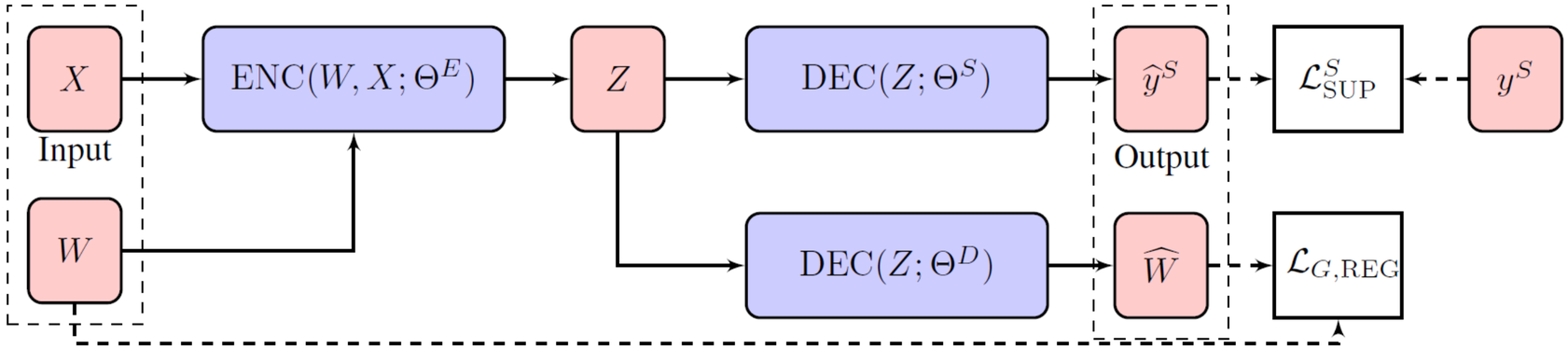
Векторизация графов (graph embeddings)

$x; (x, x')$ – данные об объектах и взаимодействиях между объектами

$z = f(x, w)$ – модель кодирования x в векторное представление z

$x' = g(z, w)$ – модель декодирования z в реконструкцию x'

$\text{Loss}(x, w) = \|x'(w) - x\| + \tau L_{\text{SUP}}(x, w_S)$ – сумма двух критериев



T.Mikolov et al. Efficient estimation of word representations in vector space, 2013.
I.Chami et al. Machine learning on graphs: a model and comprehensive taxonomy. 2020.

Перенос обучения (transfer learning)

$f(x, w)$ – часть модели, универсальная для широкого класса задач

$g(x, w')$ – часть модели, специфичная для своей задачи

$\min\limits_{w,\, w'}: \sum_x \text{Loss}_1(f(x, w), g_1(x, w'))$ – обучение по большим данным

$\min\limits_{w'} \sum_{x'} \text{Loss}_2(f(x', w), g_2(x', w'))$ – обучение по своим данным

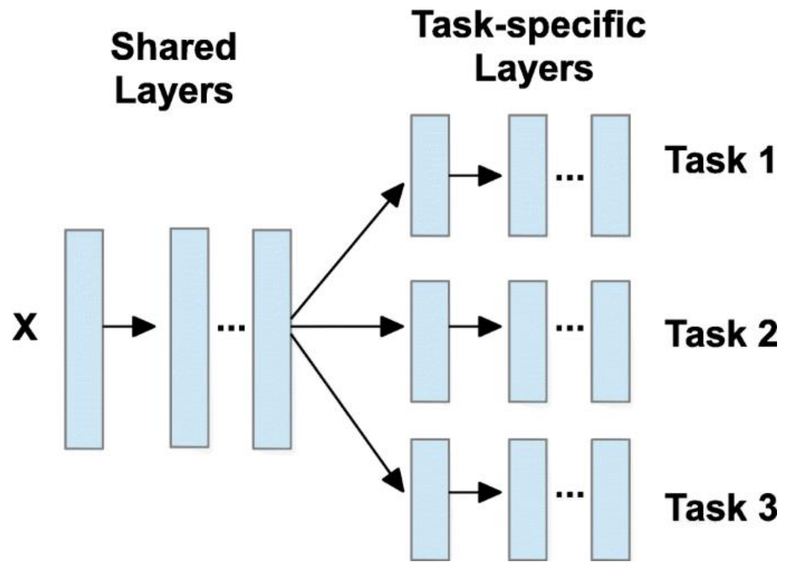
Sinno Jialin Pan, Qiang Yang.
A Survey on Transfer Learning.
2009

Многозадачное обучение (multi-task learning)

$f(x, w)$ – модель векторизации, универсальная для всех задач

$g_t(x, w'_t)$ – часть модели, специфичная для t-й задачи

$\min\limits_{w, w'_t}: \sum_t \sum_x \text{Loss}_t(f(x, w), g_t(x, w'_t))$ – обучение по всем задачам



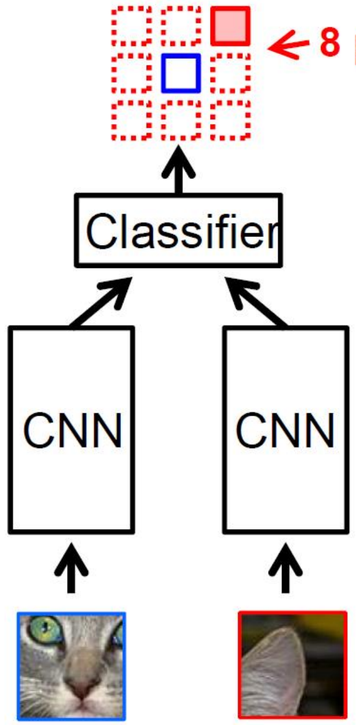
M.Crawshaw. Multi-task learning with deep neural networks: a survey. 2020

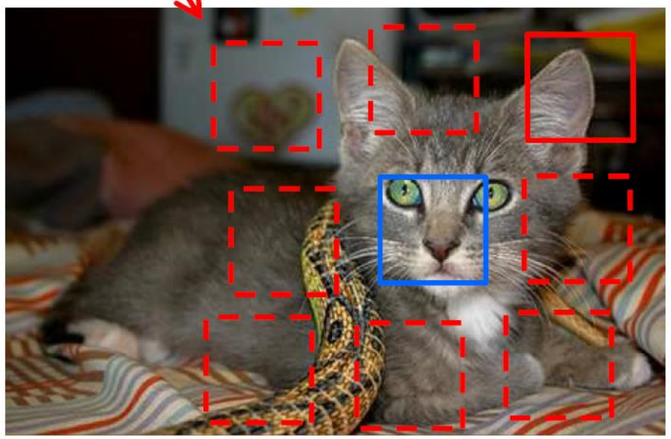
Y.Wang et al. Generalizing from a few examples: a survey on few-shot learning. 2020

Самостоятельное обучение (self-supervised)

Модель векторизации $z = f(x, w)$ обучается предсказывать взаимное расположение пар фрагментов одного изображения

Преимущество:
сеть выучивает векторные представления объектов без размеченной обучающей выборки



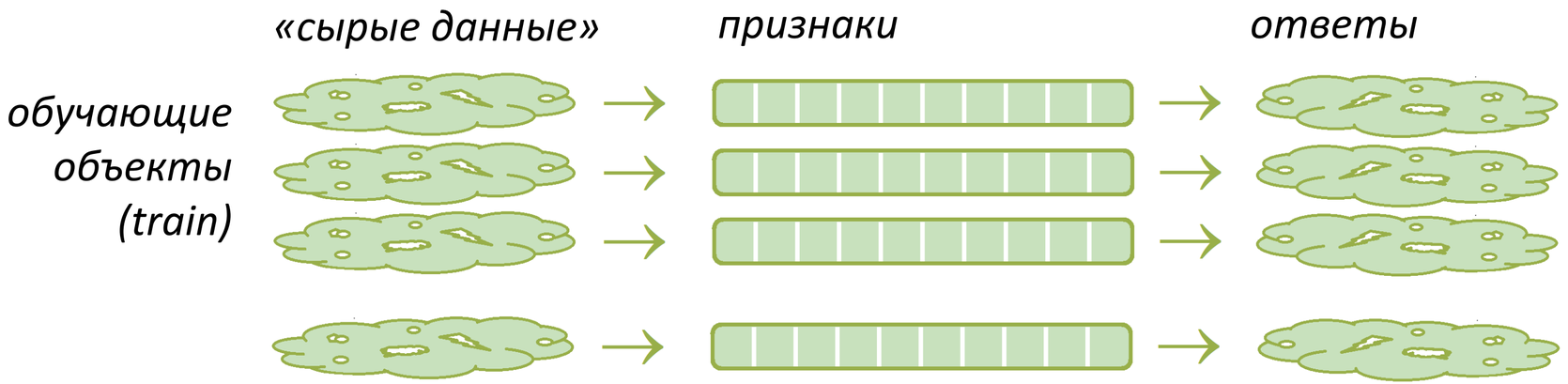


Unsupervised visual representation learning by context prediction, Carl Doersch, Abhinav Gupta, Alexei A. Efros, ICCV 2015

Нейронные сети для синтеза объектов

Вход: сложно структурированные объекты

Выход: сложно структурированные ответы



Примеры: синтез изображений, перенос стиля, машинный перевод, суммаризация текстов

Модели: seq2seq, CNN, RNN, LSTM, GAN, BERT, GPT-3 и др.

Генеративная состязательная сеть (GAN)

$x = g(z, w)$ – модель генерации реалистичного объекты x из шума z

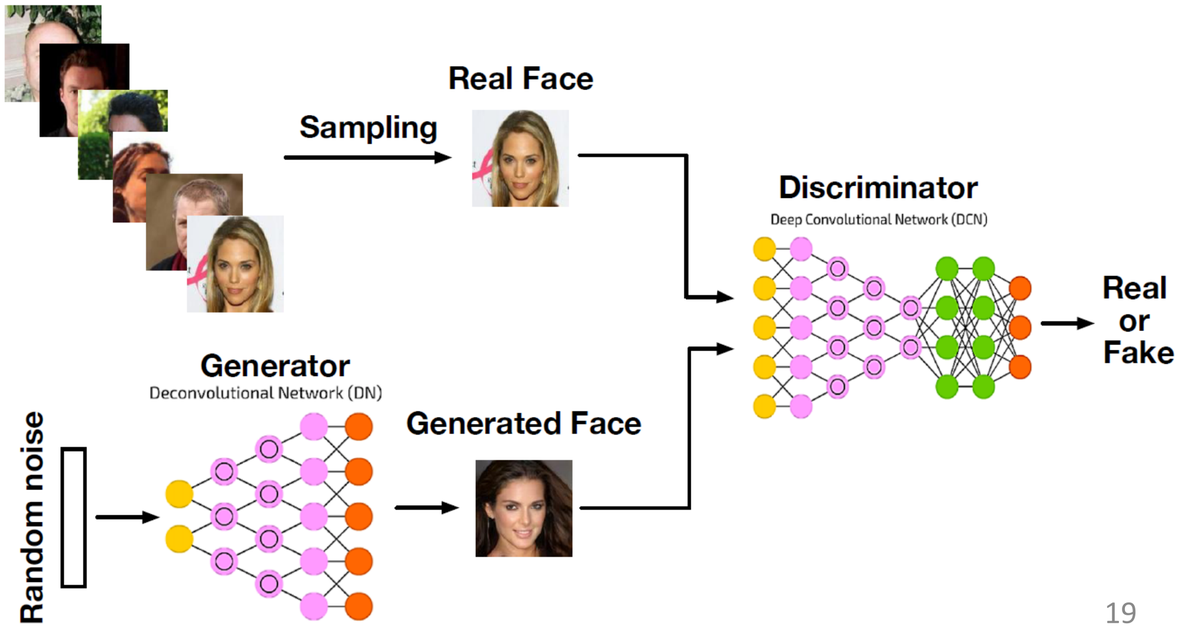
$f(x, w')$ – модель классификации x «реальный/сгенерированный»

$\min_{w} \max_{w'} \sum_x \ln f(x, w') + \ln\left(1 - f(g(z, w), w')\right)$ – совместное обучение

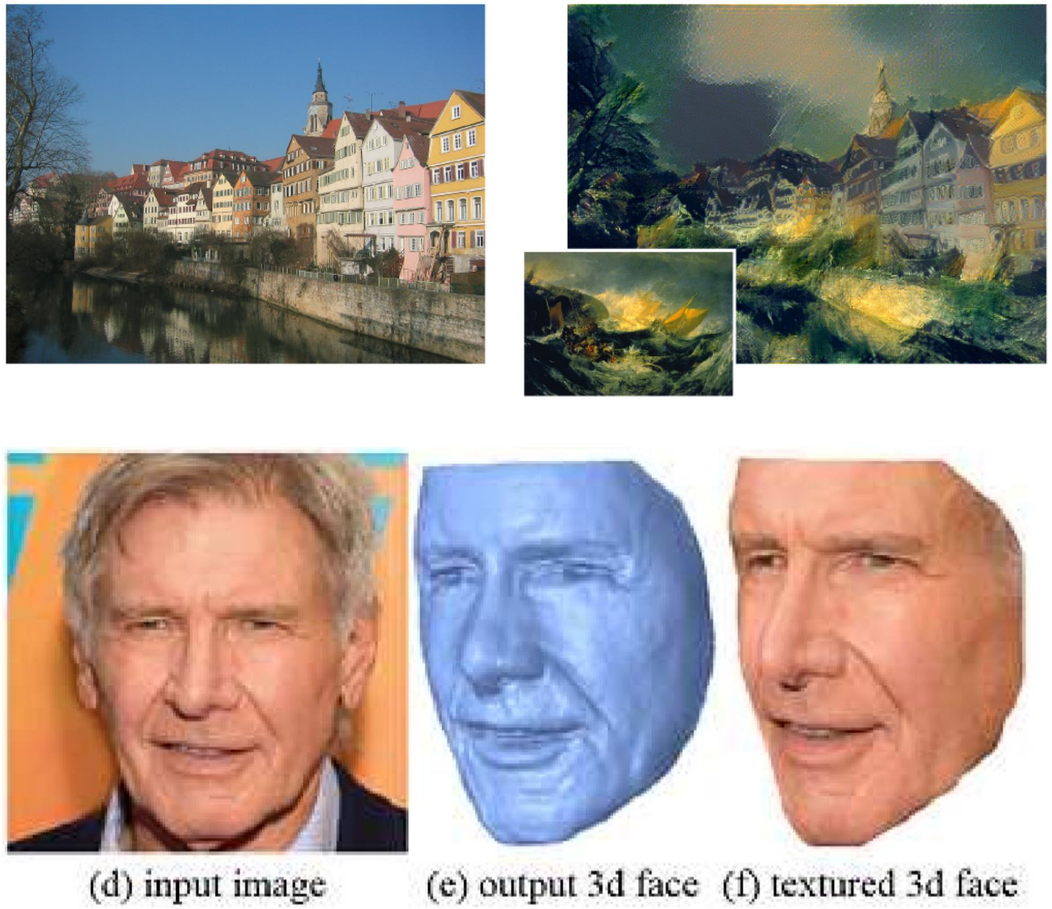
Antonia Creswell et al. Generative Adversarial Networks: an overview. 2017.

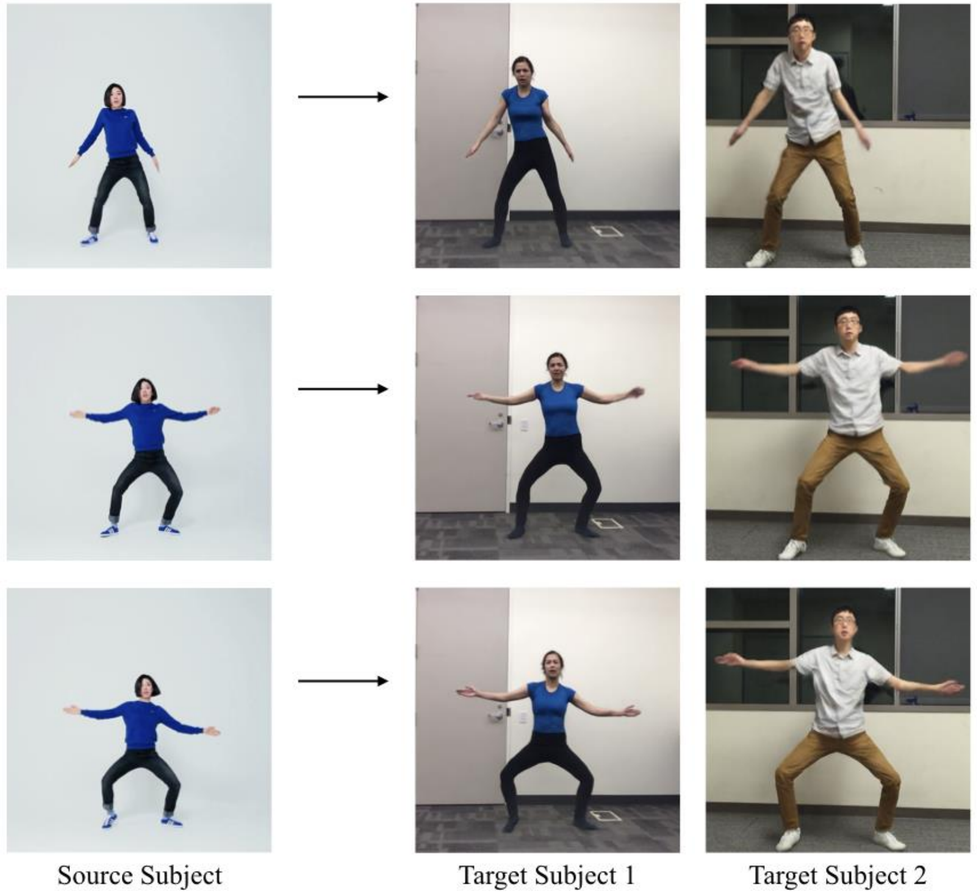
Zhengwei Wang et al. Generative Adversarial Networks: a survey and taxonomy. 2019.

Chris Nicholson. A Beginner's Guide to Generative Adversarial Networks. 2019.



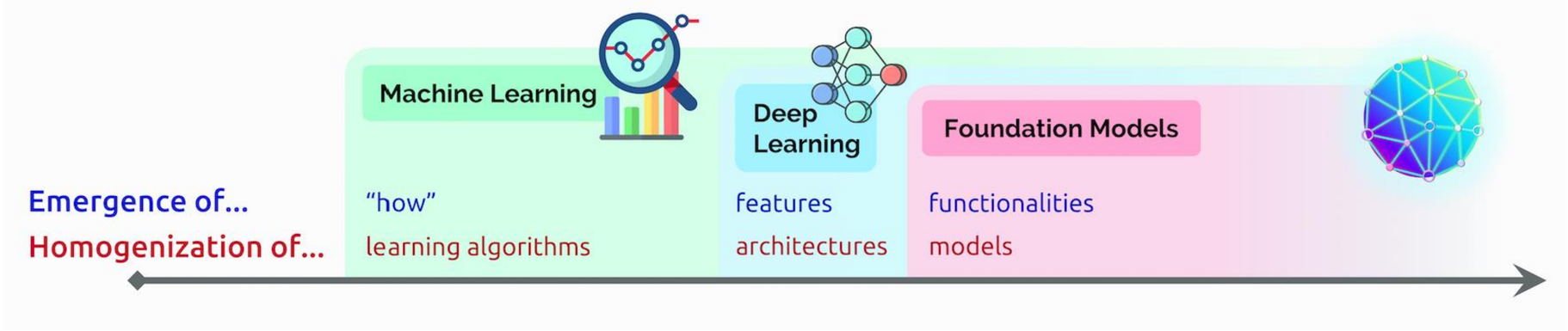
Синтез изображений и видео

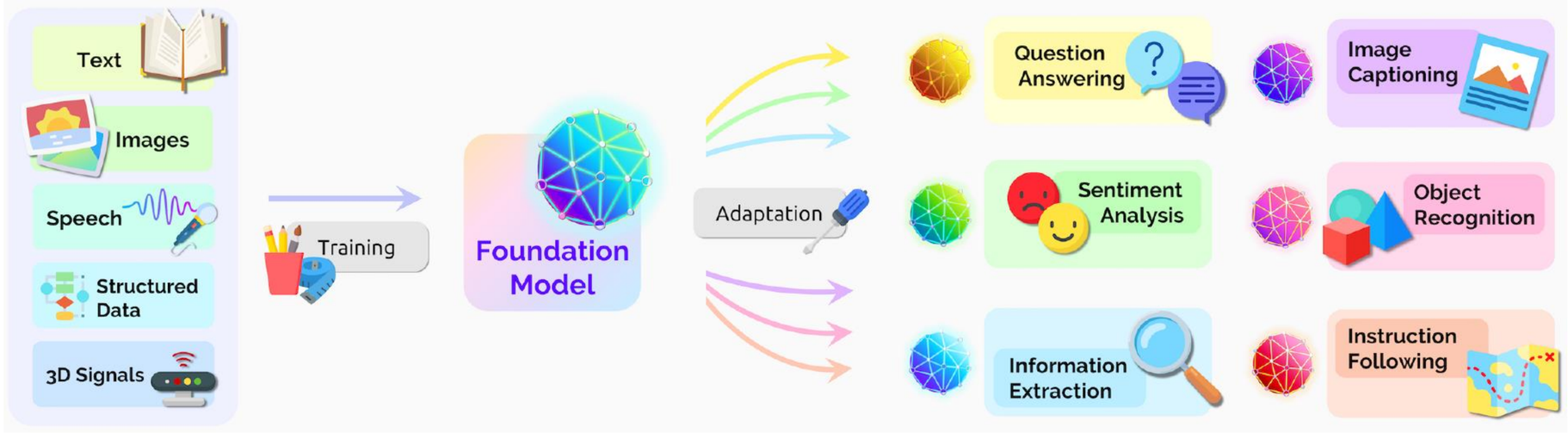




Caroline Chan, Shiry Ginosar, Tinghui Zhou, Alexei A. Efros. Everybody Dance Now. ICCV-2019.

Фундаментальные модели (Foundation Models)





R.Bommasani et al. (CRFM, Stanford University). On the opportunities and risks of foundation models. 2021

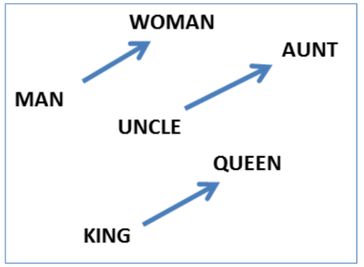
Эволюция подходов в обработке текстов

Декомпозиция задач по уровням «пирамиды NLP»

- морфологический анализ, лемматизация, опечатки, …
- синтаксический анализ, выделение терминов, NER, …
- семантический анализ, выделение фактов, тем, …

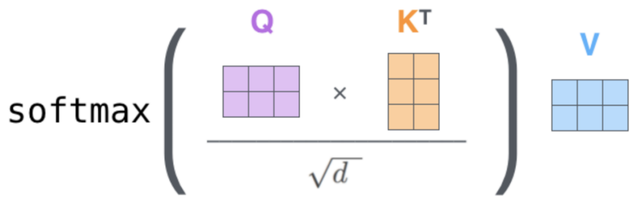
Модели векторизации слов (эмбедингов)

- модели дистрибутивной семантики:
  word2vec [Mikolov, 2013], FastText [Bojanowski, 2016], …
- тематические модели LDA [Blei, 2003], ARTM [2014], …

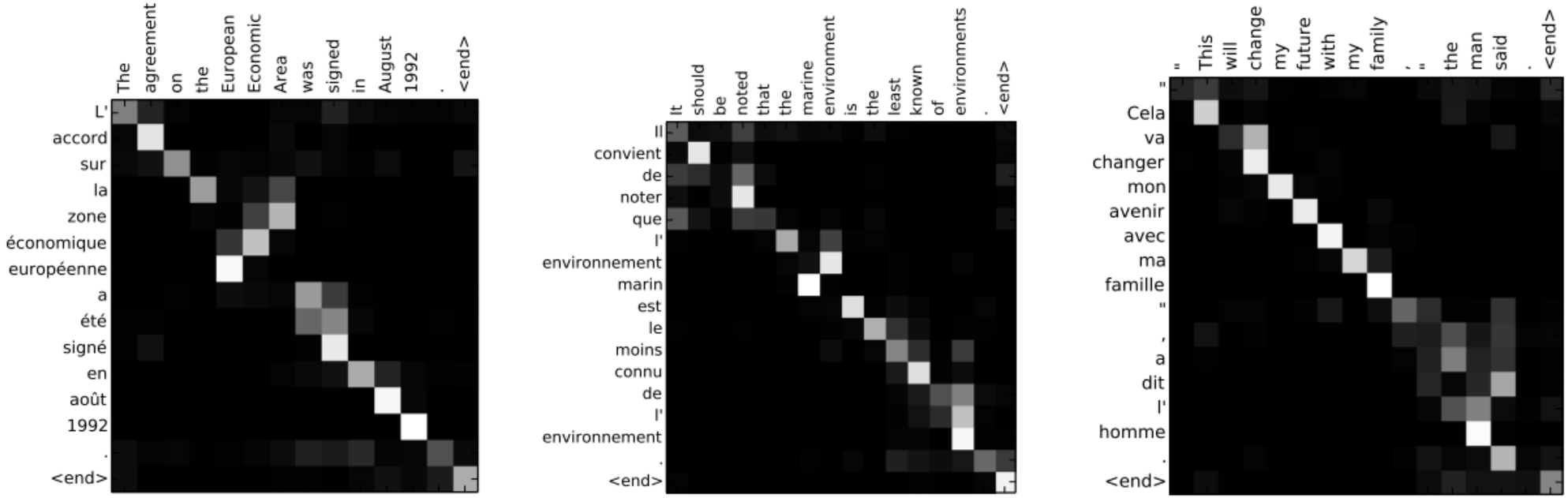


Нейросетевые модели контекстной векторизации

- рекуррентные нейронные сети: LSTM, GRU, …
- «end-to-end» модели внимания и трансформеры:
  машинный перевод [2017], BERT [2018], GPT-3 [2020], …



Модели внимания: машинный перевод



Интерпретация моделей внимания: *матрица семантического сходства* A[t,i] показывает, на какие слова x[i] входного текста модель обращает внимание, когда генерирует слово перевода y[t]

Bahdanau et al. Neural machine translation by jointly learning to align and translate. 2015

Модели внимания: аннотирование изображений

A woman is throwing a frisbee in a park.

A dog is standing on a hardwood floor.

A stop sign is on a road with a mountain in the background.

A little girl sitting on a bed with a teddy bear.

A group of people sitting on a boat in the water.

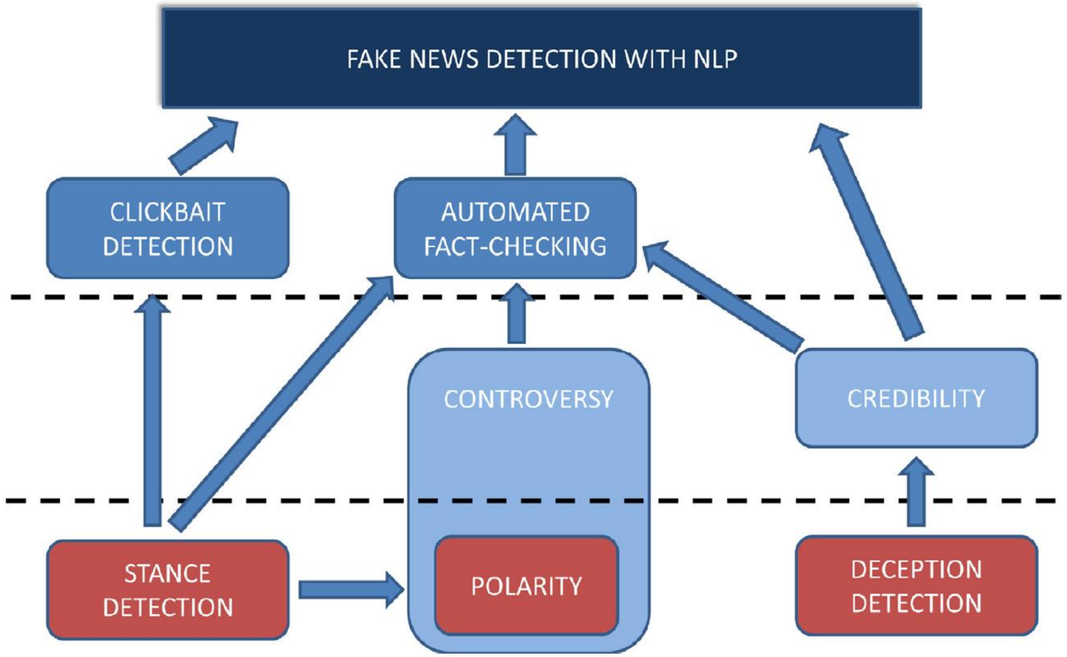
A giraffe standing in a forest with trees in the background.

Интерпретация: на какие области модель обращает внимание, генерируя подчёркнутое слово в описании изображения

Kelvin Xu et al. Show, attend and tell: neural image caption generation with visual attention. 2016

Область исследований «Fake News Detection»

1. Deception Detection
   выявление обмана в тексте новости
2. Automated Fact-Checking
   автоматическая проверка фактов
3. Stance Detection
   выявление позиции за/против запроса (claim)
4. Controversy Detection
   выявление и кластеризация разногласий
5. Polarization Detection
   классификация позиций по многим темам
6. Clickbait Detection
   выявление противоречий заголовка и текста
7. Credibility Scores
   оценка достоверности источника или новости



E.Saquete, D.Tomás, P.Moreda, P.Martínez-Barco, M.Palomar. Fighting post-truth using natural language processing: A review and open challenges. Expert Systems With Applications, Elsevier, 2020.

1. Deception Detection (выявление обмана)

- **История:** более 50 лет исследований в психологии и криминологии
- **Задача** классификации текста на два класса: *обман / не обман*
- **Обучающие выборки:**
  - Контролируемый эксперимент: люди *врут / не врут* на заданную тему
  - Материалы судебных заседаний (датасет DECOUR)
  - Отзывы на товары/услуги, проверяемые с помощью краудсорсинга
- **Признаки** – лингвистические маркеры (Linguistic-Based Cues, LBC)
- **Критерии:** Accuracy или F-мера 70–92% в зависимости от задачи
- На небольших датасетах классический ML лучше и проще DL
- Проблема переноса моделей на другие датасеты

Типы лингвистических маркеров

Манипулятивные и суггестивные приёмы

- многословие: плеоназмы, лишние слова, тавтологии, расщепления сказуемого
- избыточные повторы слов и фраз
- повышенная когнитивная сложность текста, перегруженные синтаксические конструкции
- повышенная экспрессивность, преобладание негативной тональности
- категоричность, психологическое давление

Уход от личной ответственности

- безличные глаголы, глаголы абстрактной семантики, модальные глаголы, объективация
- неконкретность, уклончивость, безличность, неопределённость высказываний

Подача информации

- оторванность от контекста: пониженная детализация места, времени, событий
- упрощение, пониженное лексическое разнообразие, лексическая недостаточность
- замалчивание фактов, сообщение ложных сведений (fact-checking, см. далее)

2. Automated Fact-Checking (проверка фактов)

- **История:** ручной fact-checking давно используется в журналистике
- **Задача** классификации текста целиком, по порядковой шкале:
  True, Mostly True, Half True, Mostly False, False
- **Обучающие выборки:**
  - Платформы для проверки фактов: Politifact, FullFact, FactCheck и др.
  - Соревнования: CLEF-2018,19,20,21, FEVER, SemEval (Rumour-Eval)
  - Датасеты: NELA-GT-2018,19, FakeNewsNet, Snopes и др.
- **Вспомогательная задача:** стоит ли отправлять текст на проверку?
  Три класса: *Non-Factual Sentence, Unimportant, Check-Worthy*
  (пример: ClaimBuster, https://idir.uta.edu/claimbuster, 2015)

3. Stance Detection (выявление позиции)

- **История:** задача textual entailment (текстового следования) – классификация пар текстов «текст *t* ⇒ гипотеза *h*» на три класса: «*h* следует из *t*», «*h* противоречит *t*», «*h* не относится к *t*»
- **Задача:** классификация текста *h* относительно запроса (claim) *t*: *agree, disagree, discusses (позиция не высказана), unrelated*
- **Обучающие выборки:**
  - SNLI: 570K пар предложений: entail, contradict, independent
  - Датасеты: Emergent, SemEval-2016 6A(stance), FakeNewsChellenge FNC-1
- **Критерии:** F1-мера до 97% на новостях; Accuracy до 68% на Twitter

4. Controversy / 5. Polarization Detection

Две специальные разновидности задачи Stance Detection

- **Controversy Detection** (выявление полемики, разногласий):
  - кластеризация мнений без учителя
  - выделение сообществ сторонников каждого мнения в социальной сети
  - количественное оценивание объёма и динамики сообществ
- **Polarization Detection** (выявление поляризованности общества):
  - выявление разногласий по совокупности запросов или тем
- **Обучающие выборки:**
  - Датасеты социальных сетей, обычно Twitter
  - Википедия
- **Критерии:** Accuracy 73–83% (на Википедии, методом kNN)

6. Clickbait Detection (обнаружение кликбейта)

- **История:** задача появилась в 2016 году. Обнаружение заголовков или ссылок-приманок, не соответствующих сути контента
- **Задача:** классификация пары «заголовок, текст» на два класса Задача аналогична Textual Entailment и Stance Detection
- **Признаки:** гиперболизация, противоречия, web-трафик
- **Обучающие выборки:**
  - Датасеты: Webis-Clickbait 2017 (32К заголовков) и др.
  - Соревнование: Clickbait challenge 2017
- **Критерии:** F1-мера до 68%; Accuracy до 86%

7. Credibility Scores (Оценивание надёжности)

- **История:** старая задача в социологии, психологии, маркетинге
- **Задача:** оценить уровень доверия (credibility, trustworthiness) для источника (СМИ, блогера, пользователя) или отдельной новости
- **Признаки:**
  - распространение ненадёжного контента (spam, deception, fake и др.)
  - вероятность быть ботом (по диспропорции рассылок и качеству контента)
  - стиль контента, геолокация и образовательный уровень читателей
- **Обучающие выборки:**
  - много несопоставимых датасетов, отсутствует «золотой стандарт»
- **Критерии:** AUC до 89%; accuracy до 81%; MSE до 0.33
  - много критериев, не хватает методологического единства

Типология потенциально опасного дискурса и система подзадач ML/NLP для его детекции

воздействия → фейки → пропаганда → инф.война

1. ■□■■ детекция приёмов манипулирования
2. ■□■■ детекция замалчивания
3. □■■■ **детекция обмана (deception detection), слухов (rumors d.), мистификаций (hoaxes d.)**
4. □■□□ **детекция кликбэйта (clickbait detection)**
5. □■□□ **автоматическая проверка фактов (auto fact-checking)**
6. □■■□ **детекция позиции (stance d.), противоречий (controversy d.), поляризации (polarization d.)**
7. □□■■ выявление конструктов картины мира: идеологем, мифологем
8. ■□■■ оценивание возможных психо-эмоциональных реакций
9. ■□■■ выявление целевых аудиторий воздействия
10. □■■■ **оценивание и предсказание скорости распространения (virality prediction)**
11. □■■□ **оценивание достоверности источников (credibility scores)**
12. □□□■ детекция прямой агрессии (угрозы, призывы, провокации, вербовка, экстремизм)

Четыре основных типа подзадач ML/NLP

1. **Классификация текста (сообщения/предложения) целиком**
   - deception detection, fact-checking, text credibility
2. **Классификация пары текстов**
   - stance, controversy, polarization, clickbait detection
   - выявление противоречий, разногласий, замалчивания
3. **Разметка текста (выделение и классификация фрагментов)**
   - поиск лингвистических маркеров (linguistic-based cues) в тексте
   - детекция приёмов манипулирования
   - выявление идеологем, ценностей, элементов социокультурного кода
   - выявление психо-эмоциональных реакций и целевых аудиторий
4. **Кластеризация или тематическое моделирование**
   - кластеризация мнений по заданной теме (controversy detection)
   - выявление поляризации общественного мнения (polarization detection)
   - выявление мнений как сочетаний фактов, тональностей, семантических ролей

Задача выявления приёмов манипулирования

Структура манипуляции:

- фрагмент-мишень
- фрагмент-воздействие
- тип манипуляции

Пример из СМИ:
«**Зеленский** просто **играет роль президента, а не является президентом**[обесценивание], – считает экс-депутат Верховной рады Борислав Береза»

Типы манипуляций (всего 18 типов):

- негативизация (обесценивание, дисфемизмы, ярлыки, депрессивы и т.п.)
- позитивизация (героизация, эвфемизация, лозунги и т.п.)
- деавторизация (замалчивание источника, маскировка под ссылку и т.п.)
- паралогизация (алогизм, ложное следование, подмена тезиса и т.п.)

Классификация приёмов манипулирования

1. Негативизация

1.1 Навешивания ярлыков
1.2 Дисфемизмы
1.3 Аналогия с негативным объектом
1.4 Антифразис
1.5 Прием обесценивания
1.6 Негативирующая гиперболизация
1.7 Моделирование негативного сценария
1.8 Вкрапление депрессивов

2. Позитивизация

2.1 Эвфемизация
2.2 Лозунговые слова и словосочетания
2.3 Позитивирующая гиперболизация

3. Деавторизация

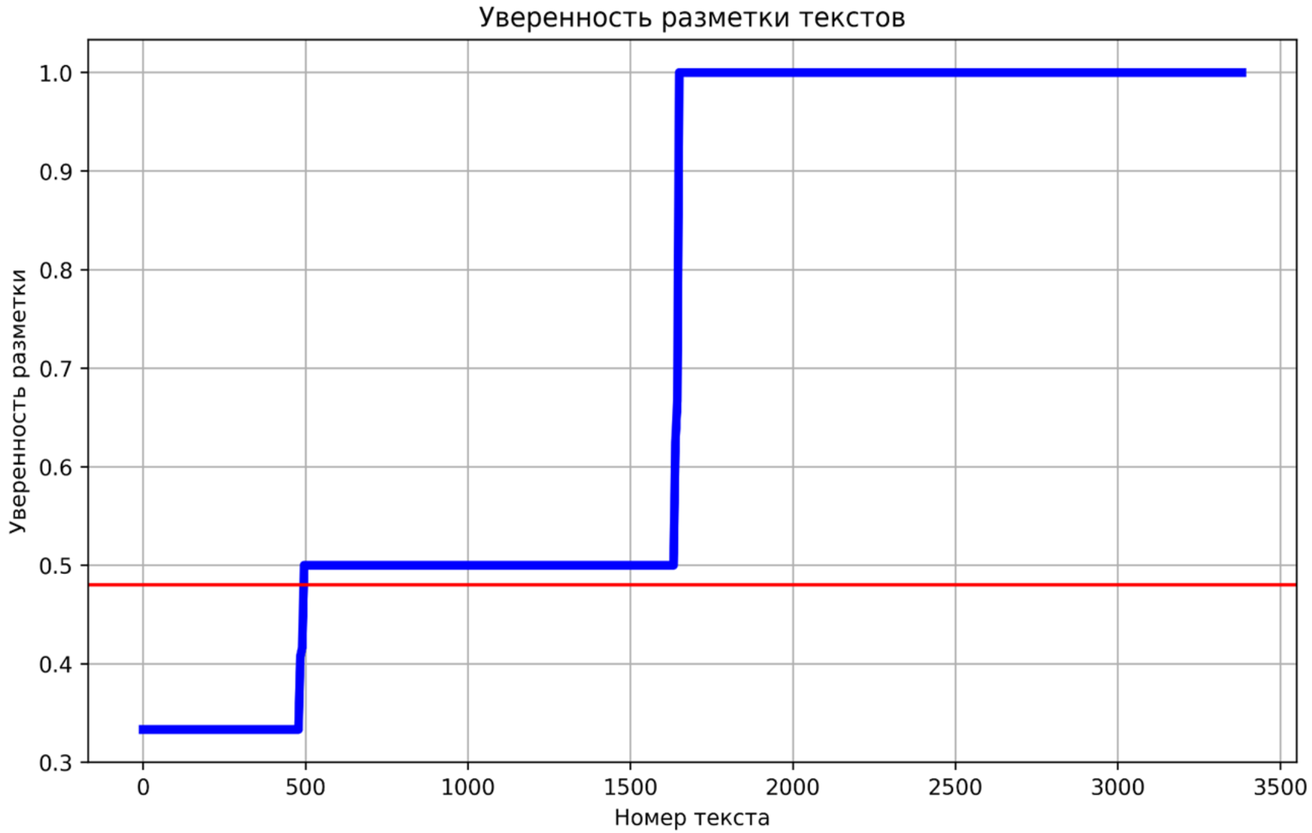
3.1 Маскировка под ссылку на авторитет
3.2 Ссылки на неопределенный источник
3.3 Ссылки на неназванных свидетелей

4. Паралогизация

4.1 Ложная причинно-следственная связь
4.2 Прием «после этого не значит поэтому»
4.3 Подмена тезиса
4.4 Высказывание о состоянии другого

Распределение размеченных данных по классам

Согласованность разметки

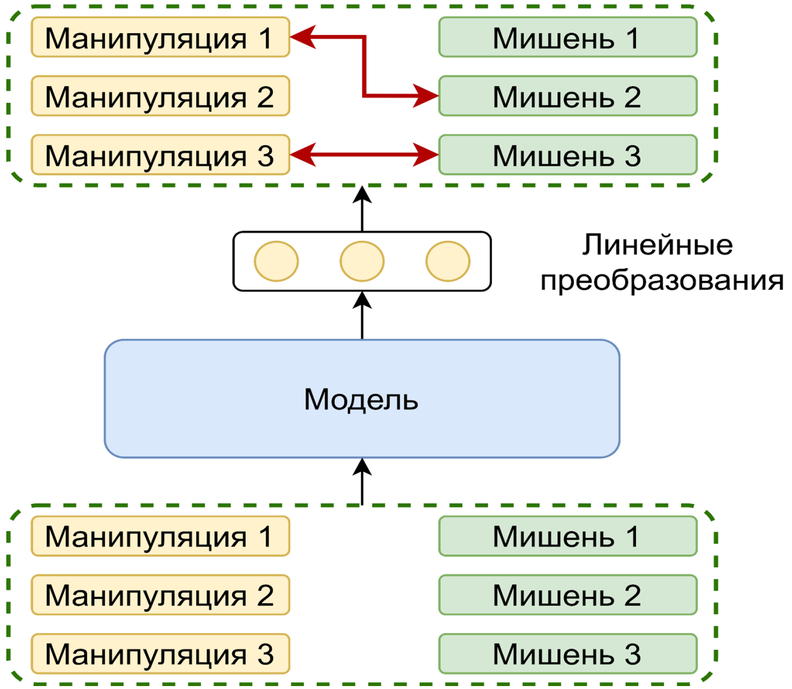


15% разметки пришлось отбросить из-за расхождения мнений разметчиков

Модели и результаты экспериментов

Двухэтапная модель:

1. Выделение фрагментов манипуляций
2. Связывание фрагментов с мишенью



	ALL	TOP-5	CATEG
JACC	0.564	0.579	0.636
$f1_{match}$	0.460	0.484	0.515
$f1_{manip}$	0.483	0.515	0.548

Таблица 3: NER_rubert + LL + RE

	ALL	TOP-5	CATEG
JACC	0.303	0.349	0.396
$f1_{match}$	0.478	0.527	0.630
$f1_{manip}$	0.223	0.241	0.301

Таблица 4: NER_xlmroberta + RE

	ALL	TOP-5	CATEG
JACC	0.544	0.557	0.604
$f1_{match}$	0.472	0.493	0.532
$f1_{manip}$	0.457	0.502	0.544

Таблица 5: NER_xlmroberta + LL + RE

	ALL	TOP-5	CATEG
JACC	0.540	0.563	0.611
$f1_{match}$	0.483	0.501	0.558
$f1_{manip}$	0.476	0.513	0.547

Таблица 6: NER + RE + LL

	ALL	TOP-5	CATEG
JACC	0.321	0.378	0.428
$f1_{match}$	0.500	0.523	0.663
$f1_{manip}$	0.252	0.295	0.323

Таблица 7: NER_rubert+ RE + LL

	ALL	TOP-5	CATEG
JACC	0.564	0.579	0.636
$f1_{match}$	0.463	0.489	0.521
$f1_{manip}$	0.483	0.515	0.548

Таблица 8: NER_rubert + LL + RE + LL

	ALL	TOP-5	CATEG
JACC	0.303	0.349	0.396
$f1_{match}$	0.481	0.513	0.624
$f1_{manip}$	0.223	0.241	0.301

Таблица 9: NER_xlmroberta + RE + LL

	ALL	TOP-5	CATEG
JACC	0.544	0.557	0.604
$f1_{match}$	0.478	0.494	0.527
$f1_{manip}$	0.457	0.502	0.544

Таблица 10: NER_xlmroberta + LL + RE + LL

Классификация приёмов пропаганды

Соревнование NLP4IF @ EMNLP-IJCNLP 2019

- 497 новостных статей из 48 источников
- 350k токенов
- 17к предложений в обучающей выборке
- Разметка outsourcing (не crowdsourcing), 6 разметчиков

Две задачи:

1. FLC (Fragment Level Classification)
   выделение и тегирование фрагментов текста
2. SLC (Sentence-Level Classification)
   бинарная классификация предложений «есть/нет манипуляция»

Da San Martino et al. Fine-grained analysis of propaganda in news articles. 2019

Соревнование NLP4IF @ EMNLP-IJCNLP 2019

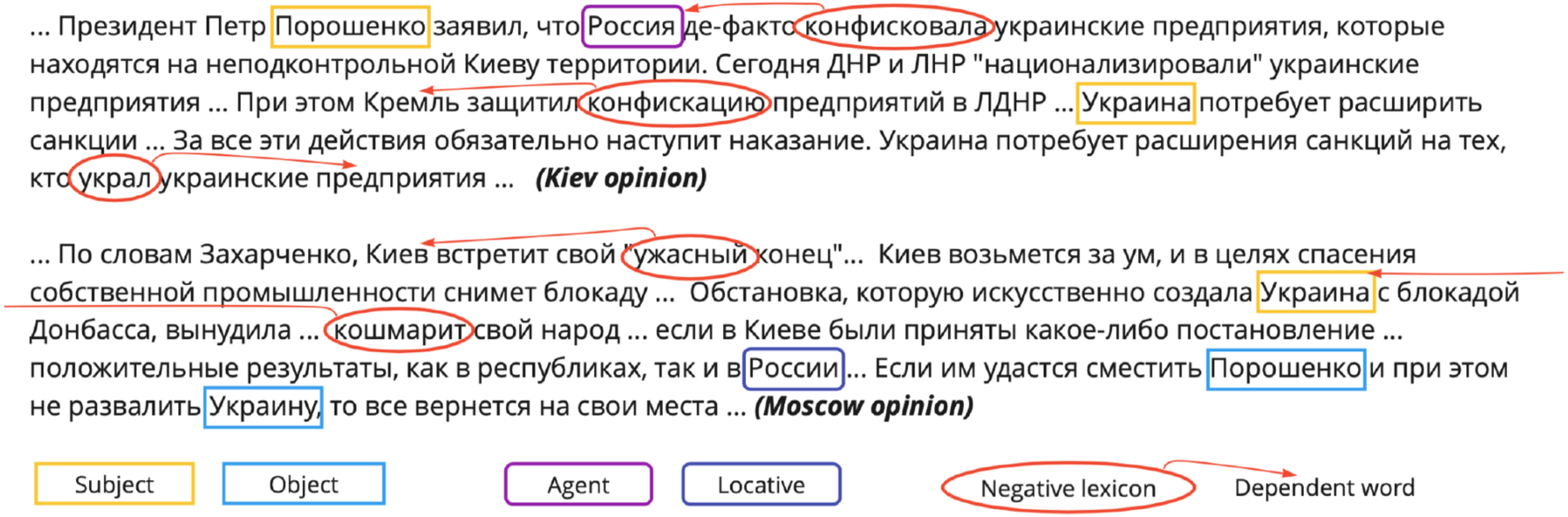
Propaganda Technique	inst	avg. length
loaded language	2,547	23.70 ± 25.30
name calling, labeling	1,294	26.10 ± 19.88
repetition	767	16.90 ± 18.92
exaggeration, minimization	571	45.36 ± 35.55
doubt	562	123.21 ± 97.65
appeal to fear/prejudice	367	93.56 ± 74.59
flag-waving	330	61.88 ± 68.61
causal oversimplification	233	121.03 ± 71.66
slogans	172	25.30 ± 13.49
appeal to authority	169	131.23 ± 123.2
black-and-white fallacy	134	98.42 ± 73.66
thought-terminating cliches	95	34.85 ± 29.28
whataboutism	76	120.93 ± 69.62
reductio ad hitlerum	66	94.58 ± 64.16
red herring	48	63.79 ± 61.63
bandwagon	17	100.29 ± 97.05
obfusc., int. vagueness, confusion	17	107.88 ± 86.74
straw man	15	79.13 ± 50.72
all	7,485	46.99 ± 61.45

- 18 приемов пропаганды
- Данных не очень много
- Сильный дисбаланс классов
- Классы очень разные по технике (loaded language & repetition)
- Классы очень разные по средней длине фрагмента

Список включает журналистские приёмы, для выделения которых не требуется внешняя информация.

Da San Martino et al. Fine-grained analysis of propaganda in news articles. (2019)

Задача выделения мнений в теме или событии



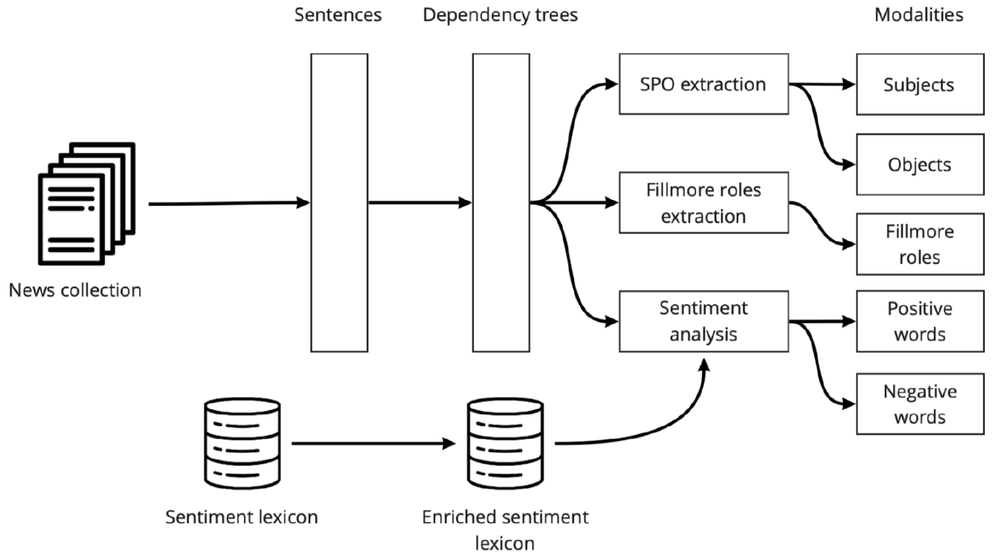
Слова «Порошенко», «Россия», «Украина» встречаются одинаково часто

«Порошенко» — субъект в первом тексте и объект во втором

«Россия» — агенс в первом тексте и локация во втором

Негативная тональность: «Россия», «Кремль» в 1-ом, «Киев», «Украина» во 2-ом

Задача выделения мнений в теме или событии



Modalities	*Pr*	*Rec*	*F*1
TF-IDF	0.51	0.95	0.67
SPO	0.59	0.7	0.64
FR	0.86	0.49	0.65
Sent	0.69	0.57	0.66
SPO+FR	0.86	0.68	0.76
SPO+Sent	0.83	0.78	0.81
FR+Sent	0.9	0.52	0.67
All	**0.77**	**0.97**	**0.86**

LPR Business

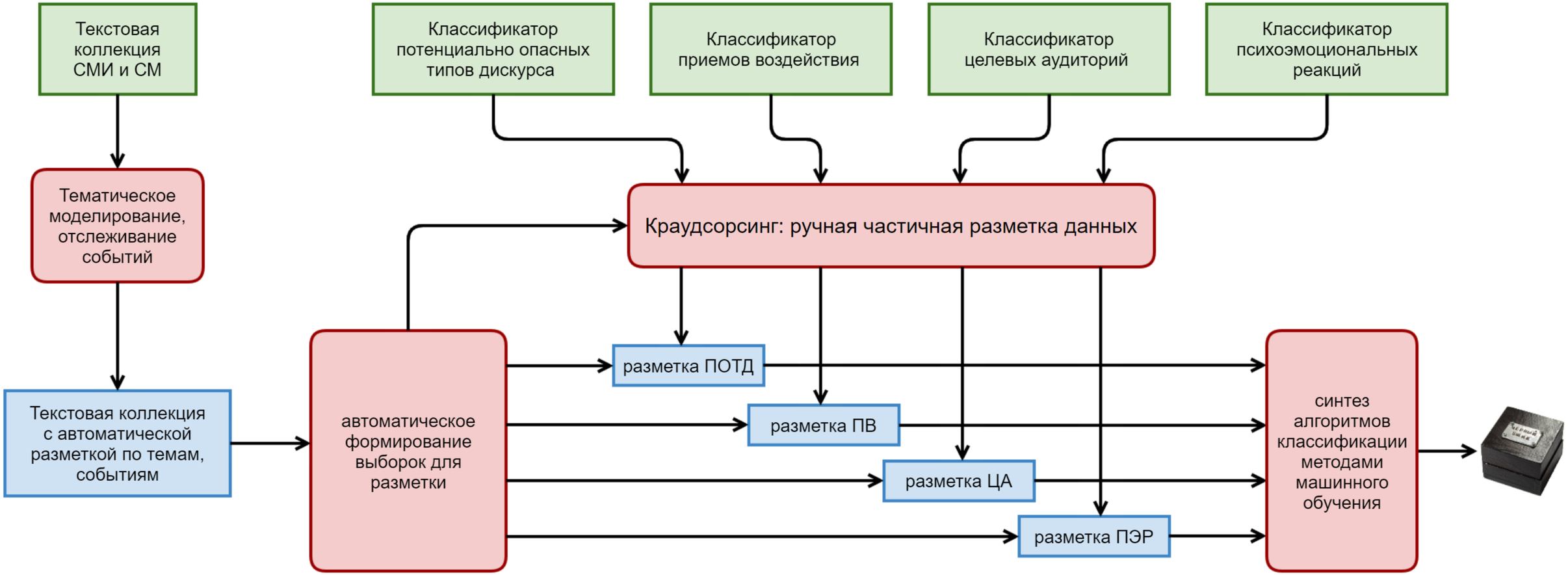
Modalities	*Pr*	*Rec*	*F*1
TF-IDF	0.57	0.97	0.72
SPO	0.56	0.99	0.72
FR	0.67	0.97	0.79
Sent	0.56	0.55	0.55
SPO+FR	0.72	0.99	0.83
SPO+Sent	0.57	0.99	0.72
FR+Sent	0.73	0.97	0.83
All	**0.77**	**0.94**	**0.85**

Paris Trump

Мнение формализуется как устойчивое сочетание слов, терминов, именованных сущностей, их семантических ролей по Филлмору и их тональных окрасок.
Все они используются в модели тематической векторизации как модальности.

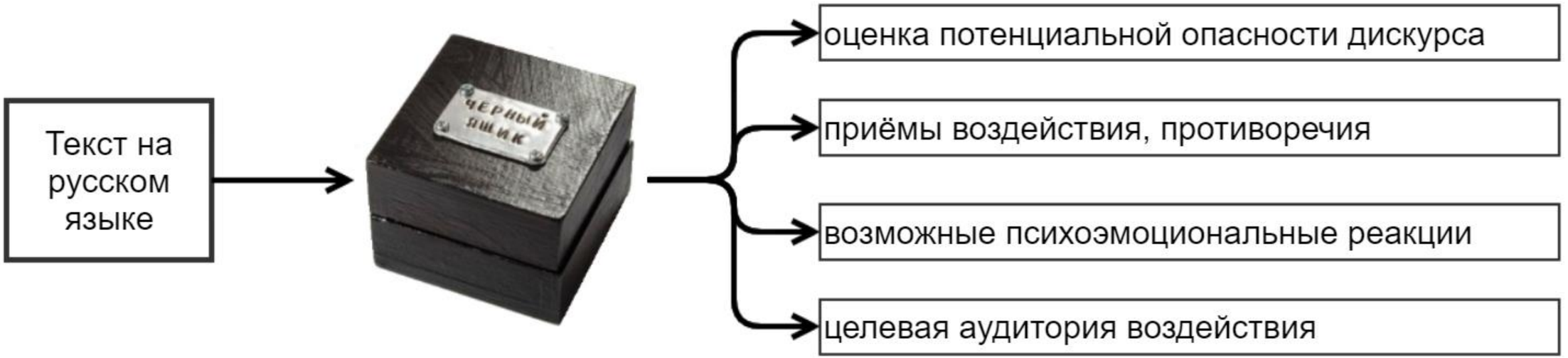
Feldman D. G., Sadekova T. R., Vorontsov K. V. Combining Facts, Semantic Roles and Sentiment Lexicon in A Generative Model for Opinion Mining. Dialogue 2020.

Разметка текстовых данных — магистральный путь формализации гуманитарных знаний



На выходе — модель классификации угроз в медийном информационном пространстве

Модель, обученная по размеченным обучающим выборкам,
может быть использована в автоматическом режиме для мониторинга
потенциально опасного дискурса в медийном информационном пространстве



КОНКУРС ПРО//ЧТЕНИЕ http://ai.upgreat.one

ЗАДАЧА:
Автоматическое выявление смысловых ошибок в текстах на естественных языках: русском и английском.

ТЕХНОЛОГИЧЕСКИЙ БАРЬЕР:
ИИ находит и аннотирует ошибки на уровне специалиста в условиях ограниченного времени.

Для конкурса собирается уникальный датасет сочинений школьников, смысловые ошибки выявляются преподавателями и экспертами ЕГЭ в соответствии со стандартами ФИПИ

ФАКТИЧЕСКАЯ ОШИБКА
автор высказывания А.Франц

В своем высказывании «Если человек зависит от природы, то и она от него зависит» Д. Мережковский говорит о необходимости защиты природы.

ЛОГИЧЕСКАЯ ОШИБКА
тезис не обоснован

ПРИЗОВОЙ ФОНД

РУССКИЙ
100 МЛН. РУБ.

АНГЛИЙСКИЙ
100 МЛН. РУБ.

Пример. «Проблема ответственности за свои поступки (по тексту В. Ф. Тендрякова)»

Алгоритмическая разметка

Нередко люди совершают плохие поступки, забывая о том, что, даже скрыв свой поступок от других, человек не скроется от своей совести. Что же такое безнравственный поступок? Безнравственный поступок - это поступок, не соответствующий моральным нормам.

Можно ли оправдать безнравственный поступок? Именно эту проблему В. Ф. Тендряков поднимает в своем тексте. Докажем сказанное примерами из представленного отрывка.

В тексте В. Ф. Тендряков говорит о том, что человек во благо себе может легко совершить низкий поступок, не испытав при этом чувство стыда. Человек сможет оправдать свой поступок перед самим собой, объяснив причину. В пример автор приводит поведение героя, который часто в жизни совершал безнравственные поступки. Он врал, дрался и крал. Мы видим, что до войны герой привык совершать плохие поступки. Он всегда оправдывался, потому что не хотел нести ответственность за свои действия, а значит не испытывал мучения совести. Мы знаем, что муки совести — это первое и самое сильное наказание, которое получает человек, совершивший плохой поступок. Но наш герой не получал никакого наказания и поэтому продолжал совершать безнравственные поступки. Проанализировав поведение главного героя, я убедилась в том, что человек обязан нести ответственность за свои поступки всегда, и поэтому я утверждаю, что нельзя оправдывать даже мелкие безнравственные поступки.

СВЯЗЬ Р.ПОВТОР
Р.ПОВТОР Р.ЛИШН ПРОБЛЕМА
Р.ПОВТОР Р.ПОВТОР Р.ПОВТОР
Р.ЛИШН
Р.ПОВТОР
Р.ПОВТОР
Р.ПОВТОР
Р.ПОВТОР Г.ОДНОР Г.ОДНОР Г.ОДН
Г.ВИДОВР Р.ПОВТОР
Р.ПОВТОР Р.ПОВТОР
Р.ПОВТОР Р.ПОВТОР
Р.ПОВТОР Г.ВИДОВР Р.ПОВТОР
Р.ПОВТОР
Р.ПОВТОР

Экспертная разметка 2

Нередко люди совершают плохие поступки, забывая о том, что, даже скрыв свой поступок от других, человек не скроется от своей совести. Что же такое безнравственный поступок? Безнравственный поступок - это поступок, не соответствующий моральным нормам.

Можно ли оправдать безнравственный поступок? Именно эту проблему В. Ф. Тендряков поднимает в своем тексте. Докажем сказанное примерами из представленного отрывка.

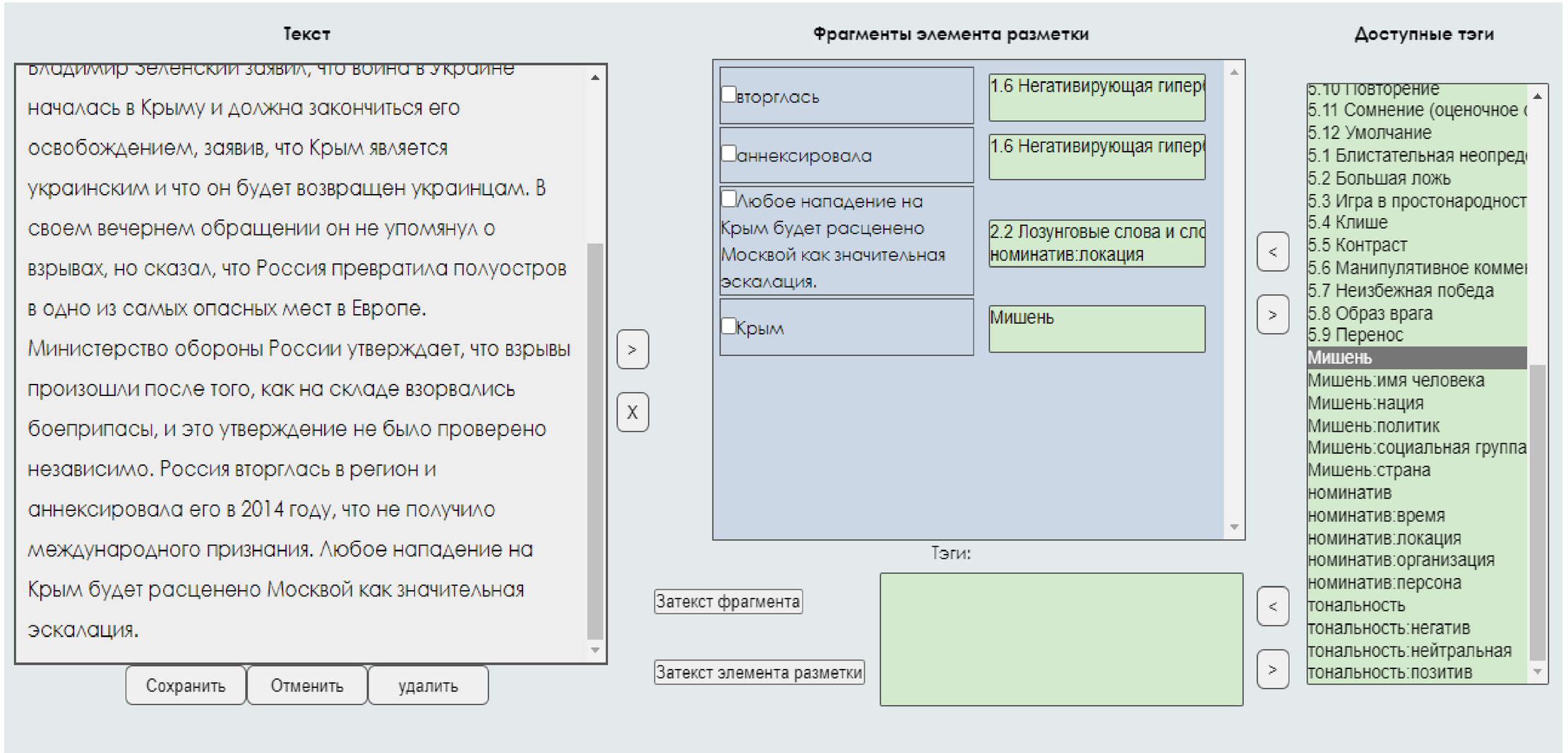
В тексте В. Ф. Тендряков говорит о том, что человек во благо себе может легко совершить низкий поступок, не испытав при этом чувство стыда. Человек сможет оправдать свой поступок перед самим собой, объяснив причину. В пример автор приводит поведение героя, который часто в жизни совершал безнравственные поступки. Он врал, дрался и крал. Мы видим, что до войны герой привык совершать плохие поступки. Он всегда оправдывался, потому что не хотел нести ответственность за свои действия, а значит не испытывал мучения совести. Мы знаем, что муки совести — это первое и самое сильное наказание, которое получает человек, совершивший плохой поступок. Но наш герой не получал никакого наказания и поэтому продолжал совершать безнравственные поступки. Проанализировав поведение главного героя, я убедилась в том, что человек обязан нести ответственность за свои поступки всегда, и поэтому я утверждаю, что нельзя оправдывать даже мелкие безнравственные поступки.

Р.ПОВТОР Т1
Р.ПОВТОР Т1
Р.ПОВТОР Т2 Р.ПОВТОР
ПРОБЛЕМА Р.ПОВТОР Т
ПРИМЕР Р.ПОВТОР Т3
Р.ТАВТ Т4 Р.ПОВТОР Т1
Р.ПОВТОР Т1
Р.ТАВТ Т4
Р.ПОВТОР Т1
Р.ТАВТ Т4 Р.ПОВТОР Т1
Р.ТАВТ Т4 Р.ПОВТОР Т1
Р.ТАВТ Т4 Р.ПОВТОР Т1
ПОЯСНЕНИЕ
Р.ПОВТОР Т1
Р.ПОВТОР Т1

Структура разметки текстовых данных

1. Документ может иметь несколько элементов разметки
2. Каждый элемент разметки может иметь
   - несколько меток
   - несколько фрагментов текста
   - несколько затекстов (текстов, добавляемых разметчиками)
3. Каждый фрагмент и затекст может иметь несколько меток
4. Словари меток (иерархии классов, категорий, тегов)
5. Словари затекстов (например, мемы, цитаты, анекдоты и т.п.)

Инструмент разметки текстовых данных



Выводы

1. Машинное обучение — это оптимизация параметров моделей
2. ИИ — не «интеллект», а обучаемая векторизация данных
3. Перспективы развития ИИ — автоматизация процесса CRISP-DM
4. Предобученные модели внимания / трансформеры позволяют теперь решать сложные задачи понимания естественного языка
5. В том числе стоить модели для мониторинга и детекции угроз в медийном информационном пространстве
6. Разметка текстовых данных — магистральный путь формализации гуманитарных знаний в таких задачах
7. Методология разметки и оценивания требует стандартизации

Спасибо за внимание!

Воронцов Константин Вячеславович
д.ф.-м.н., профессор РАН,
руководитель лаборатории МОСА
(Машинного Обучения и Семантического Анализа)
Института Искусственного Интеллекта МГУ
voron@mlsa-iai.ru

http://www.MachineLearning.ru/wiki?title=User:Vokov