Проекты направления

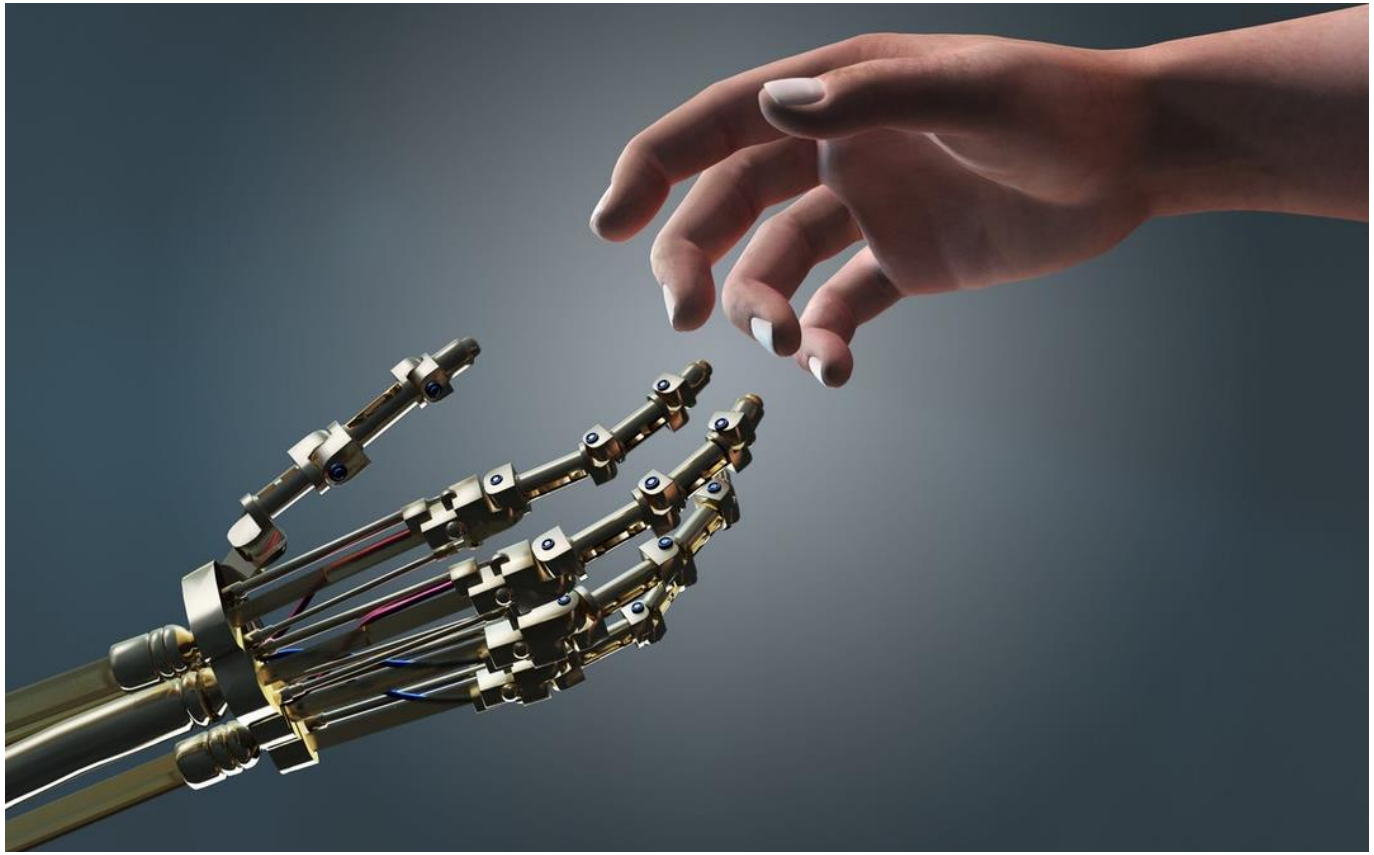
«Машинное обучение и анализ данных»

Руководитель направления – Шпильман Алексей Александрович

alexey@shpilman.com

Human AI Interaction

In collaboration with SimLabs



Daniel Kudenko

daniel.kudenko@york.ac.uk

Motivation

AI is becoming an everyday presence for the average person: **human-AI interaction.**

Increased network connectivity (e.g. internet of things) enables collection of behaviour data: **data analytics & machine learning.**


Image: Reuters

Research Vision

- Intersection of human-AI interaction with data analytics leads to many research challenges and opportunities in a wide range of applications.

- Overall Goals:
  - Design adaptive AI systems that efficiently collaborate with users and support them in their tasks.
  - Use human behavioural data to train and improve AI systems.

Research Questions

- What are the types of support that a user will need when interacting and collaborating with an AI system?
- How can we recognize the support need of human users?
- How can the AI system accurately predict human behaviour?
- How accurate does this prediction have to be?
- How can control sharing be effectively realized in an Human/AI controlled system?
- How is control handover best realized in shared-control systems? How can human and AI agents be jointly trained to learn optimal cooperative behaviour?
- How can human/AI interaction be optimized in more complex multi-agent systems with multiple human and AI agents?
- How can data from human behaviour be used to design or improve AI behaviour?
- Many more questions to be discovered on the way ……..

Examples of Current Application Areas

- Software Development:
  - Intelligent programmer assistance.
- Air traffic control:
  - Controller training support.
  - Pilot behaviour modelling.
- MOOCs:
  - Learner performance prediction.
  - Personalized exercise generation.
- Robotics
  - Socially aware robots

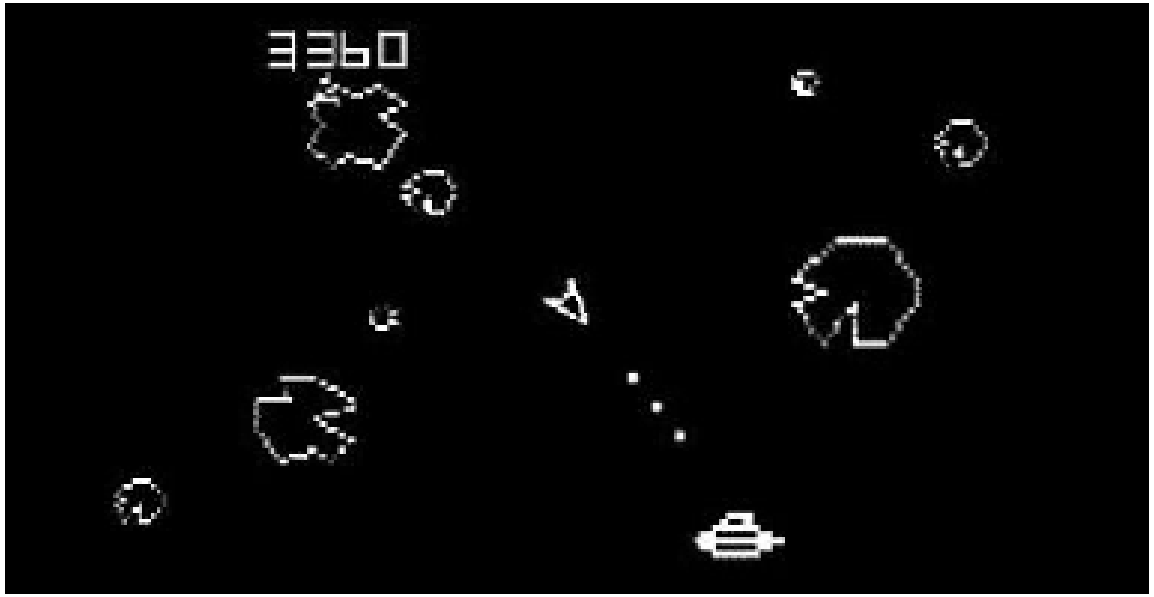
Human-AI Shared Control

Human Controller

AI Controller

Asteroids Project

- Implement (or use an implementation of) an action game, e.g. Asteroids, where there are different aspects to control, e.g. moving and shooting.
- Train a reinforcement learning agent to move, shoot, and do both.
- Compare performance of three settings:
  - AI shooting and human moving.
  - AI moving and human shooting.
  - Human controlling everything.
  - AI controlling everything.



- How can we train the AI to optimally support the human?
- Can a human behavioural model be used for this training?
- Training schedules for shared control.
- Ideal (but not reserved) for Masters student considering to continue with a PhD.

Requirements: familiarity with RL techniques (optimal)

1 student

Collaboration Setting

Human Builder

AI Builder

Minecraft Projet

- Experiment domain: Microsoft Project Malmo (Minecraft).
- Decide on some distributed task (e.g. building a castle).
- Train a reinforcement learning agent to support a human player in that task.
- Evaluation criteria: performance gain when
  - Human performing the task on their own.
  - Human receiving assistance from RL agent.

Requirements: familiarity with RL techniques (optimal)

1 student

Crowd-sourced AI

- Lots of data on human decision making becoming available.

- Rather than creating an AI from scratch, use crowd-sourced data to compute a decision in a given situation.

Crowd CS-GO Project

- Data source: computer game, e.g. CS-GO (extracting data from game recordings).
- Alternative: simulate crowd by implementing multiple stochastic heuristic AIs.
- Given a game situation, use crowd data to compute a decision.
- Ideas:
  - Weighted majority voting (e.g. by expertise).
  - Averaging.
  - State similarity matching.

  Requirements: willingness to work with weird interface of CS data

  1 student

Taking Game AI to the next level

- 1997: Deep Blue vs Kasparov: Chess AI beats human champion
- Minimax search with expert heuristic and optimization/parallelization.

- 2016: AlphaGo vs : Chess AI beats Go champion.
- Monte Carlo Search with Deep Learning.

What’s next?

- Go: number of possible actions per move and state space is big, but not at “real-world” level.

- Hex-and-counter games raise complexity to a new level.

Vassal Project

- Vassal is an open source GUI for hex-and-counter games (and others).
- Only used to make and send moves, but does not check legality of moves and does not have an AI.
- First stage: create AI API for Vassal.
- Second stage: implement the AlphaGo approach and test it on a hex-and-counter game.
- Third stage: take game AI to the next level!

Requirements: familiarity with a DL library

1 student

Machine Learning Methods in Software Engineering Research Group

Тимофей Брыксин

timofey.bryksin@gmail.com

JET BRAINS RESEARCH

Факторизация Github’a

- Выделение стиля написания кода программистов с Github.
- Поиск программистов с похожим стилем.
- Проверка авторства кода.
- Система анти-плагиата для домашних задач по программированию.

Требования: Python или что угодно другое

1-2 человека

Snippet search

- Анализ в реальном времени создаваемого кода
- Предложение программисту вариантов решения его текущей задачи
  (от шаблонов и сниппетов до готовых кусков кода)
- https://www.codota.com/, только более умная и удобная

Требования: Java/Kotlin

1 человек

Анализ задач для Stepik

- Определение сложности задачи по тексту её условия и коду решения.
- Генерация текстов условий и кода решения задач по желаемой теме, уровню сложности и т.п.

Требования: Python или что угодно другое.

1-2 человека.

Machine Learning and Information Management Lab

Игорь Кураленок

ikuralenok@gmail.com

JET BRAINS RESEARCH

Яндекс

Поиск №1 в России*

Шаблоны поиска

При поиске часто возникает желание получить ответ определенного шаблонного вида, например по товару хочется цену, отзывы, фотки и характеристики, по фильму рейтинги, билеты, отзывы трейлер и т.п..

Более того, по каждой части мы примерно понимаем чего хотим и зачастую можем описать ход поиска той или иной части.

Цель работы систематизировать эти знания и разработать язык шаблонизации поиска на котором подобные поисковые сценарии задания вопросов и извлечения ответов можно было бы описывать.

Требования: пройти собеседование у Игоря Кураленка.

1 студент.

Обучение с подкреплением roguelike

Есть известная игра https://crawl.develz.org, хочется сделать робота, который умеет в нее играть.

Основное отличие от того, что делали в частности в deep mind, в том, что моделирование самой задачи хочется делать автоматически и то, что игра длится довольно долго, так что невозможно получить миллионы прецедентов.

Требования: пройти собеседование у Игоря Кураленка.

1 студент.

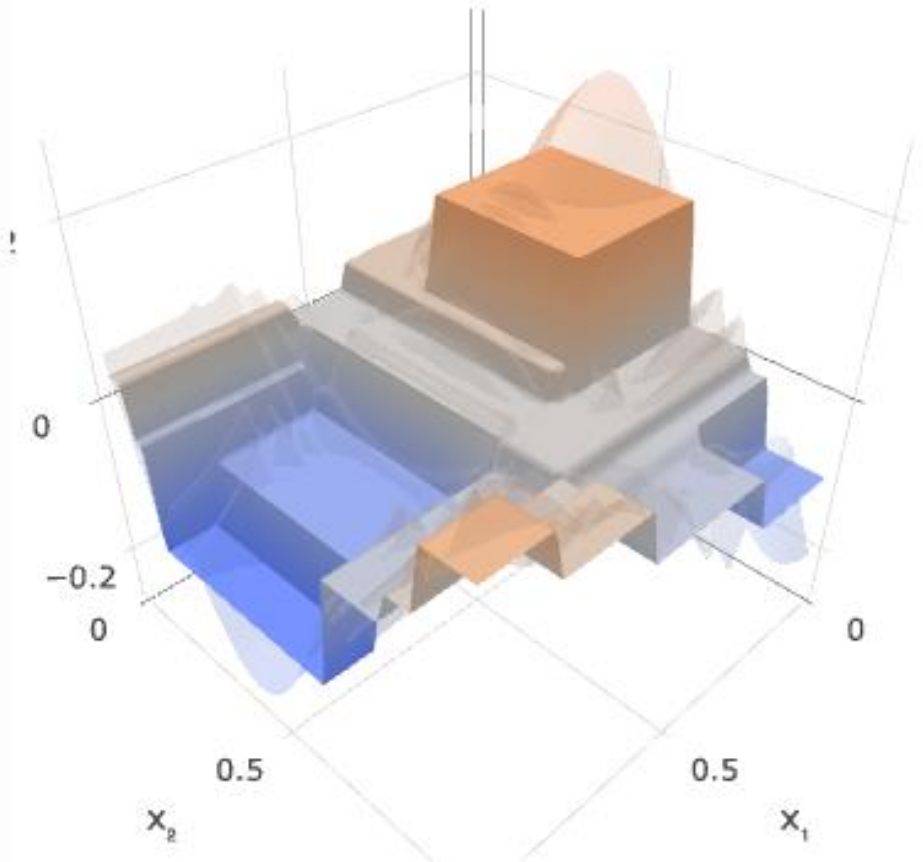
Feature extraction
в процессе градиентного бустинга

Мы все любим нейронные сети за то, что они умеют решать одновременно задачу оптимизации и задачу выделения новых свойств из существующих данных, полезных для решения поставленной задачи.

Хочется научиться решать подобную задачу и для градиентного бустинга, где также есть подобная возможность.

Требования: пройти собеседование у Игоря Кураленка.

1 студент.

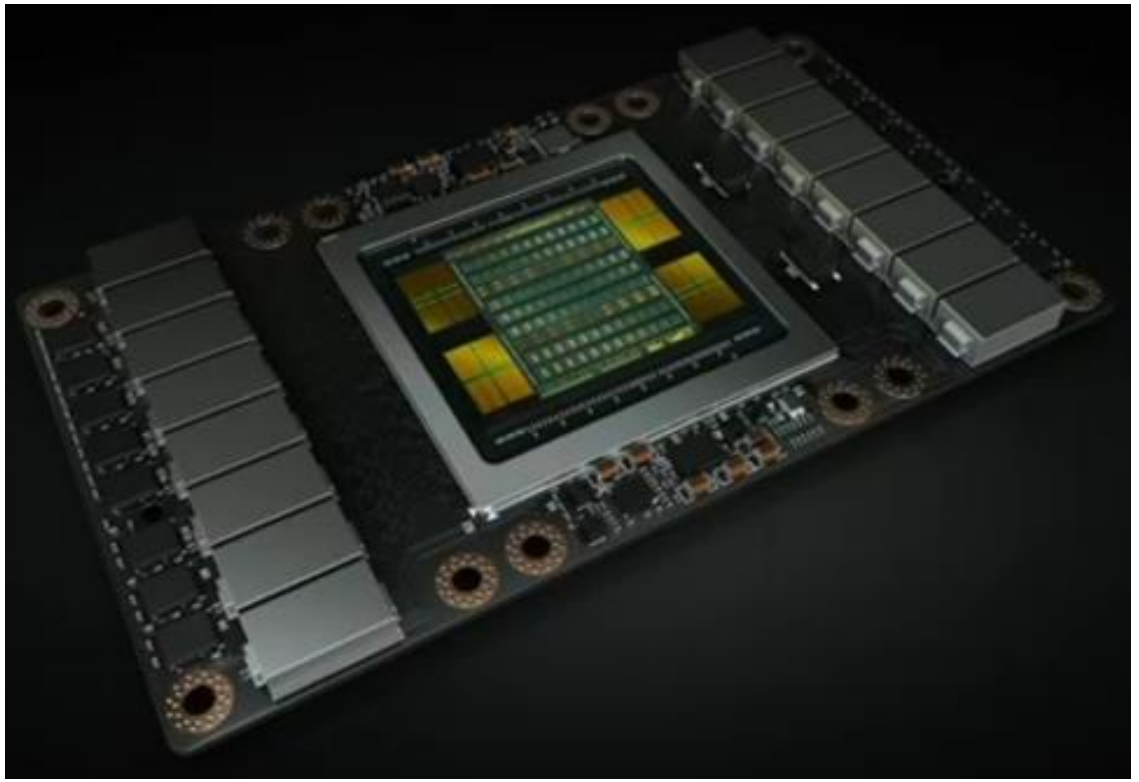


Реализация GBDT на GPU

Есть деревья, их можно очень быстро обучать на GPU, хочется это реализовать. Задача очень практическая, реализовывать будем на C++ CUDA и python.

Требования: хороший уровень C++.

1 студент.



Байесовские модели в рекомендациях

Проблема построения хороших рекомендаций очень актуальна.

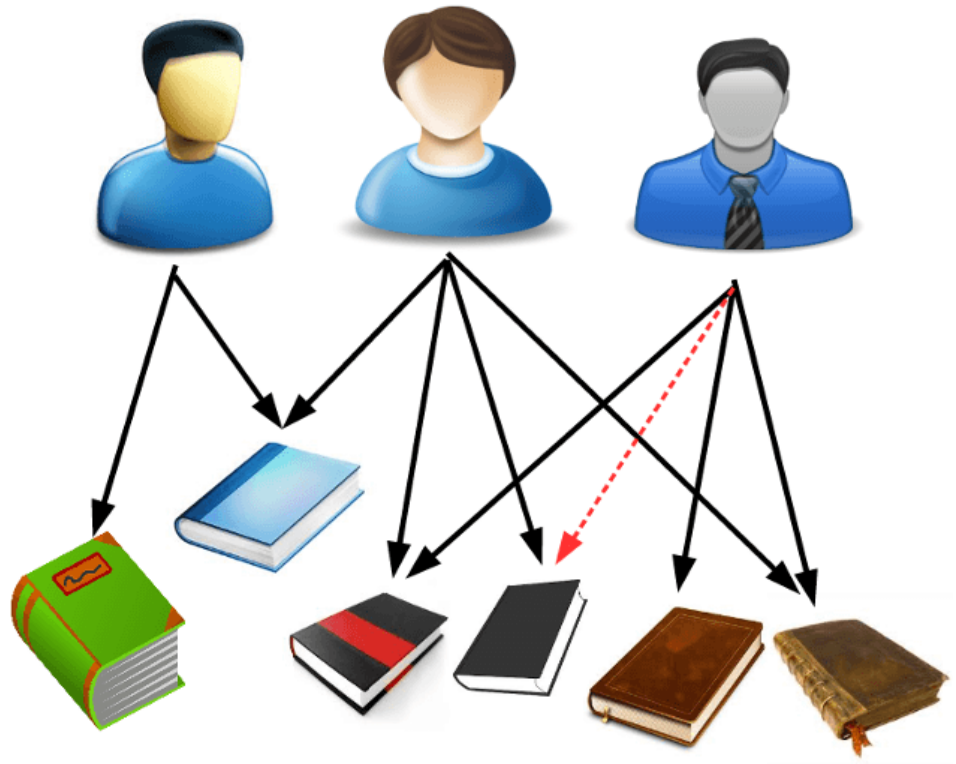
Однако хороших способов строить эти рекомендации пока не проглядывается.

Хочется попробовать несколько иной подход к их построению на основе байесовских моделей.

В частности будем больше уделять внимания не столько предсказанию оценки товара, сколько предсказанию того какой товар будет оценен следующим.

Требования: пройти собеседование у Игоря Кураленка.

1 студент.



Term sharded поиск по википедии

Существующие поисковые системы построены по принципу шардирования по данным.

Следуя этому принципу надо разложить порции инвертированного списка на сервера постранично, вычислить ранжирующую функцию, передать результаты для выбора лучших вверх по иерархии.

Такой подход прост в реализации ранжирующей функции, но неэффективен по распределению нагрузки: на каждый запрос, не попавший в кеш мы запрашиваем все машинки.

Существует альтернативный подход, так называемый term sharded, при котором локальным является индекс по одному терму, а вычисление ранжирующей функции происходит выше.

Было несколько попыток реализации данного подхода (например Microsoft Maguro/Tiger), которые неудались из-за недостаточного учета связей между разными терминами при фильтрации.

Есть идея как эту проблему излечить с помощью умного алгоритма выделения n-gramm.

Требования: пройти собеседование у Игоря Кураленка.

1 студент.

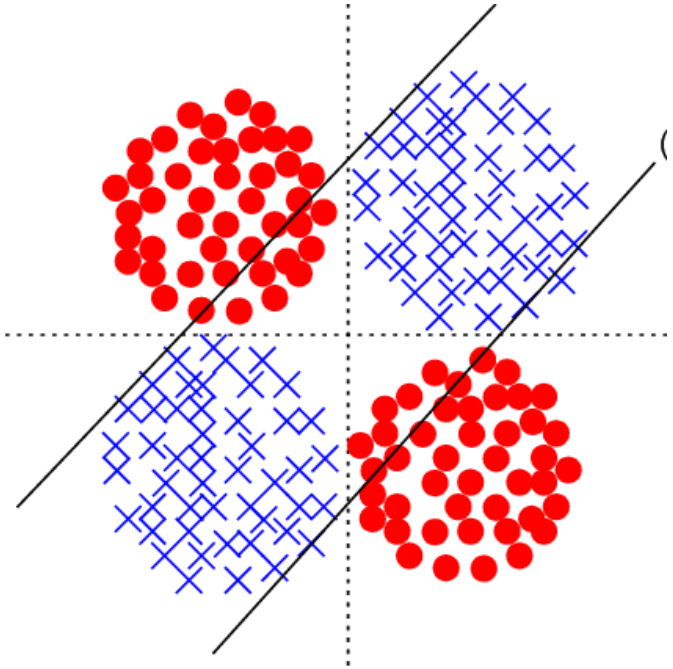
Жадные регионы vs. деревья

Структура дерева для некоторых подходов излишне сложна, в частности, она предполагает биение всего пространства жесткими границами.

Есть способ подбирать не целое дерево а регион, задаваемый предикатом, основанном на свойствах точки, и константы или распределения, ассоциированный с подходящими по условию точками.

При этом способ не проигрывает существующим деревьям при объединении в ансамбль.

Хочется описать существующий метод, показать, что он работает на реальных данных и показать, что данная конструкция лучше приспособлена к использованию категориальных признаков.



Требования: пройти собеседование у Игоря Кураленка.

1 студент.

Общие проекты направления

Шпильман Алексей Александрович

alexey@shpilman.com

Мультицелевой генетический алгоритм для оптимизации модели клеточного пространства

Одной из основных проблем в моделировании сложных систем является подборка параметров.

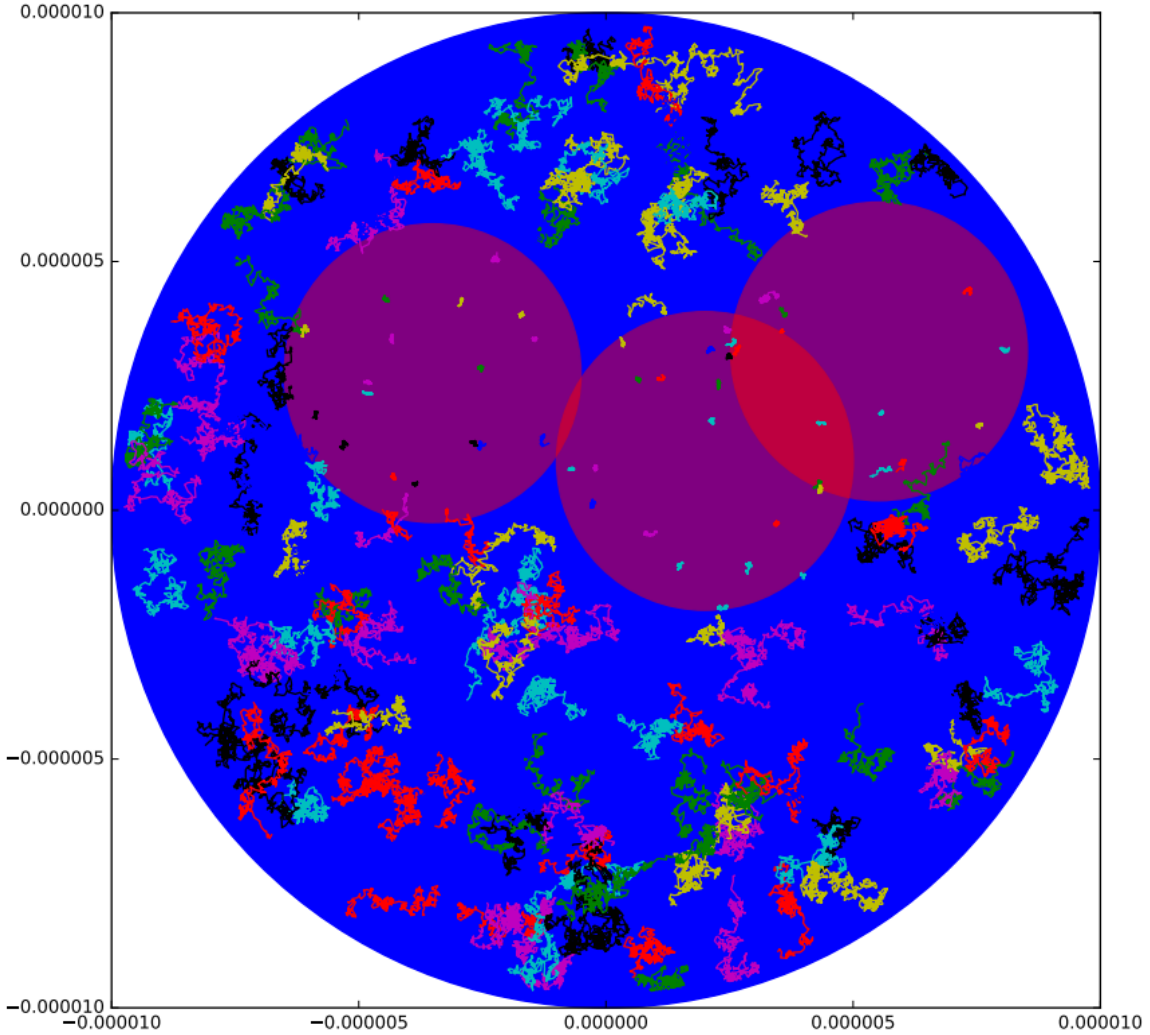
В последнее время для решения таких задач активно используется машинное обучение, в частности алгоритмы стохастической оптимизации, например генетический алгоритм.

Дополнительной сложностью служит отсутствие единой метрики соответствия модели реальности.

Работа состоит в реализации генетического алгоритма для приближения модели по нескольким целевым функциям сразу.

Требования: Python, Numpy, Matplotlib.

1-3 человека.



Legal AI

В данный момент совместно с юристами мы разрабатываем Deep Learning систему анализа ключевых мест в судебном решении.

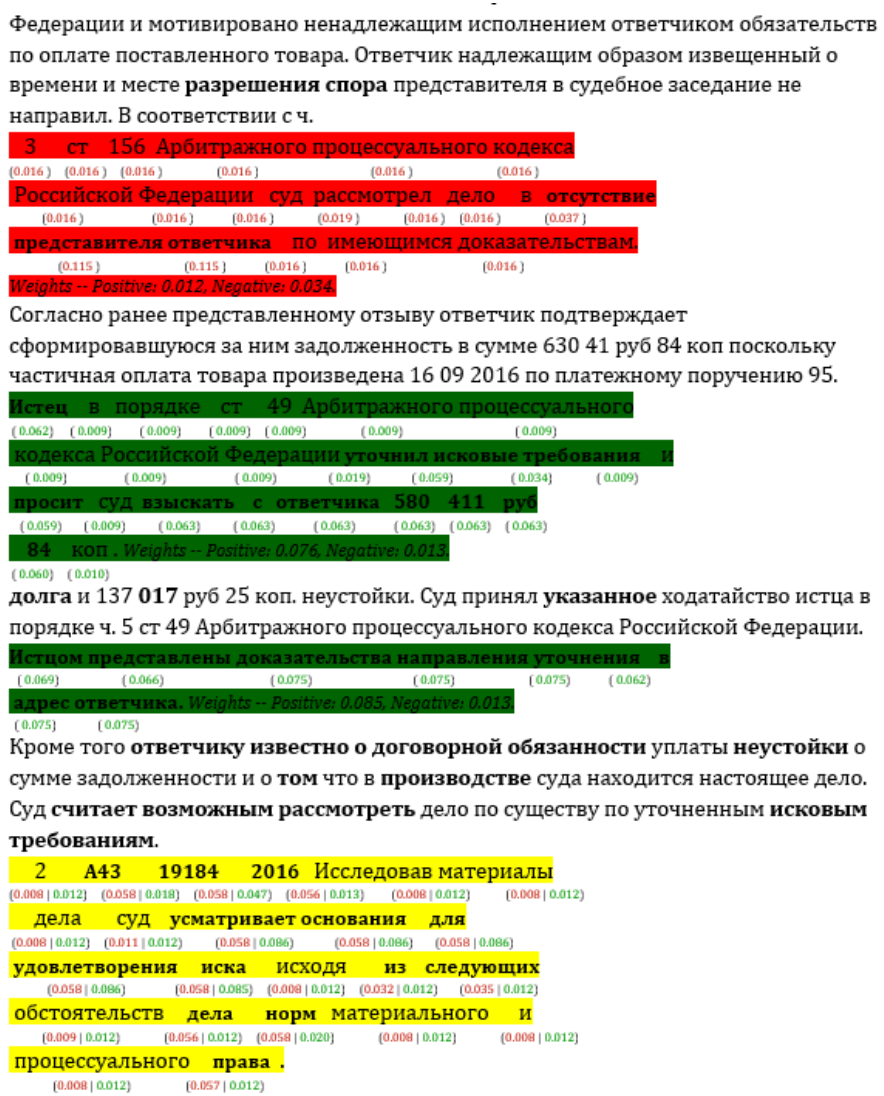
Проект будет посвящен ка разработке инструментов для потенциальных пользователей на основе этой системы и усовершенствованию системы как таковой.

Непосредственная задача будет состоять в выделении ключевых факторов

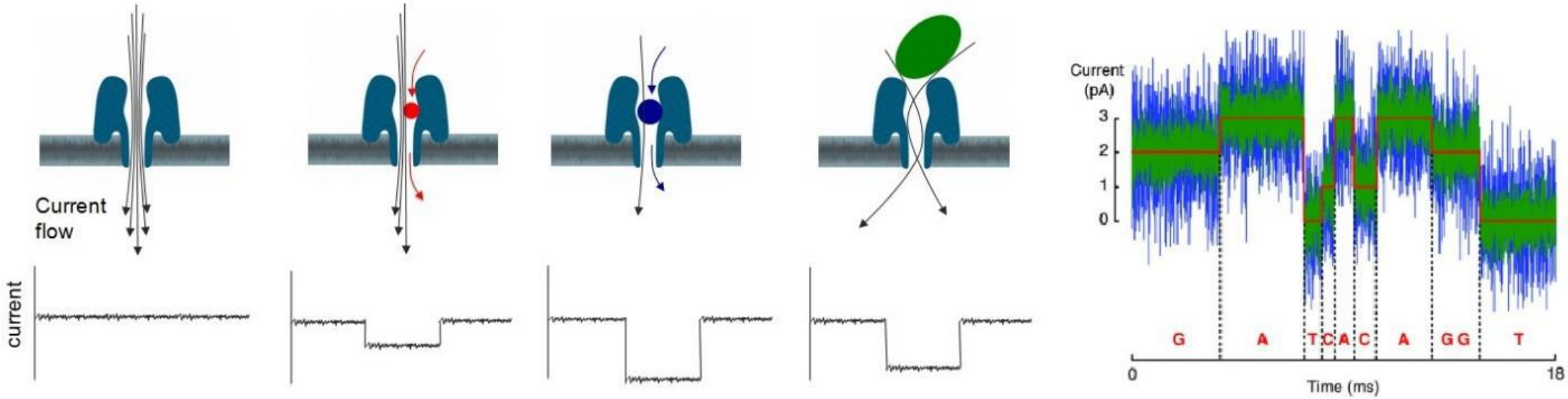
Планируется применить подход active learning.

Требования: Python, Pandas, Test (PDF) mining, Tensorflow*.

1 человек.



Deep Learning for Nanopore Sequencing



Next-next-generation sequencing. После пропускания через нанопору нити ДНК мы получаем сигнал, который можем переводить в последовательность.

Существует несколько нерешенных проблем, которые пока мешают достигнуть необходимой точности для массового использования.

Требования: Any DL library.

1 человек.