КУМСКОВ МИХАИЛ ИВАНОВИЧ

KUMSKOV@MAIL.RU

MIKHAIL.KUMSKOV@MATH.MSU.RU

СЕМИНАР

«МЕТОДЫ АНАЛИЗА МОЛЕКУЛЯРНЫХ ГРАФОВ В ЗАДАЧЕ «СТРУКТУРА-СВОЙСТВО»

СТРУКТУРНЫЕ ОБЪЕКТЫ

- ***Направление*** — задачи классификации и прогнозирования на объектах, которые можно представить простыми помеченными графами (на Структурных Объектах).

ЧТО ЭТО ЗА ОБЪЕКТЫ?

- Молекулярные графы
- Изображения
- Временные ряды (*финансовых котировок*)

МОЛЕКУЛЯРНЫЕ ГРАФЫ

Задача прогнозирования физико-химических свойств М-графов.

QSPR – Quantitative Structure-Property Relationship

Задача прогнозирования лекарственной и биологической активности М-графов

QSAR - Quantitative Structure-Activity Relationship

ИЗОБРАЖЕНИЯ

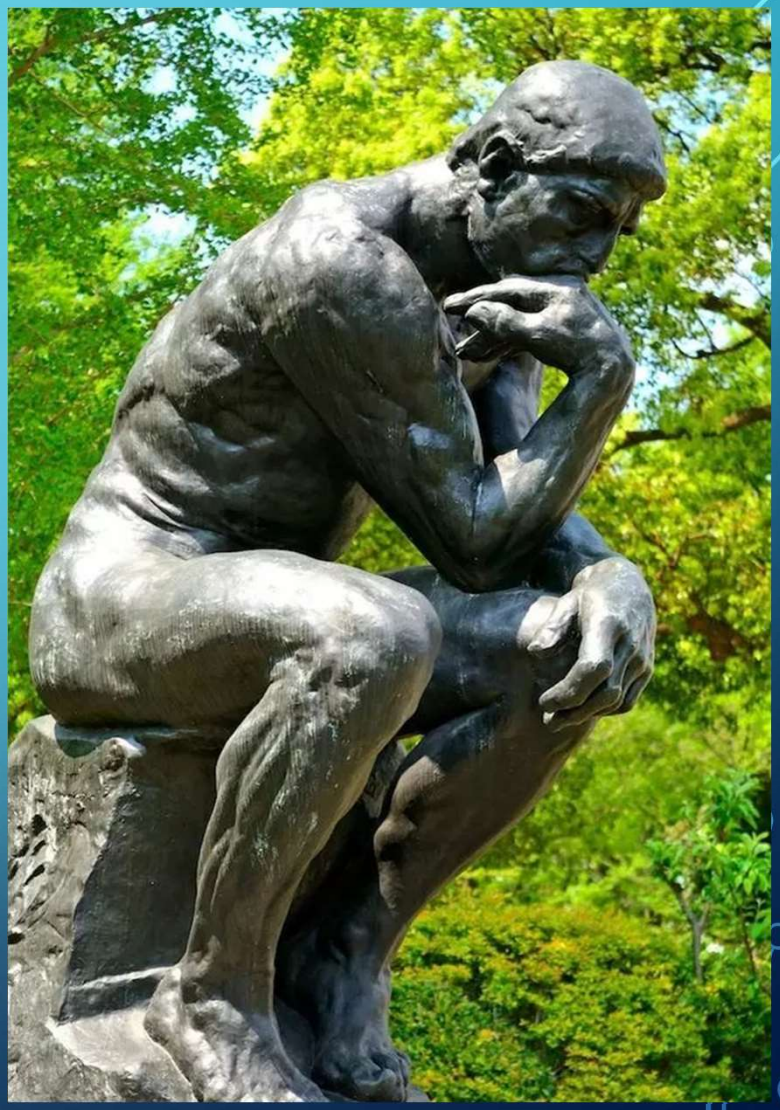
- Задача поиск ***заданных объектов*** на картинке
- Задача восстановление ***позы человека***-спортсмена
- Задача ***сегментации*** «плоских» изображений – поиск точек описания (Особых Точек – ОТ)
- Задача *сегментации* изображений на *поверхности М-графа*

ВРЕМЕННЫЕ РЯДЫ

- Задача прогнозирования *трендов* финансовых котировок
- Задача написание торговых *ботов*, где реализующих модели прогноза тренда котировок

НАПРАВЛЕНИЕ ИССЛЕДОВАНИЙ

- Найти и оптимизировать «правильное» **представление** Структурного Объекта в виде простого помеченного графа
- Конструируем разметку вершин графа так, чтобы минимизировать ошибку прогноза в ***заданном классификаторе***.

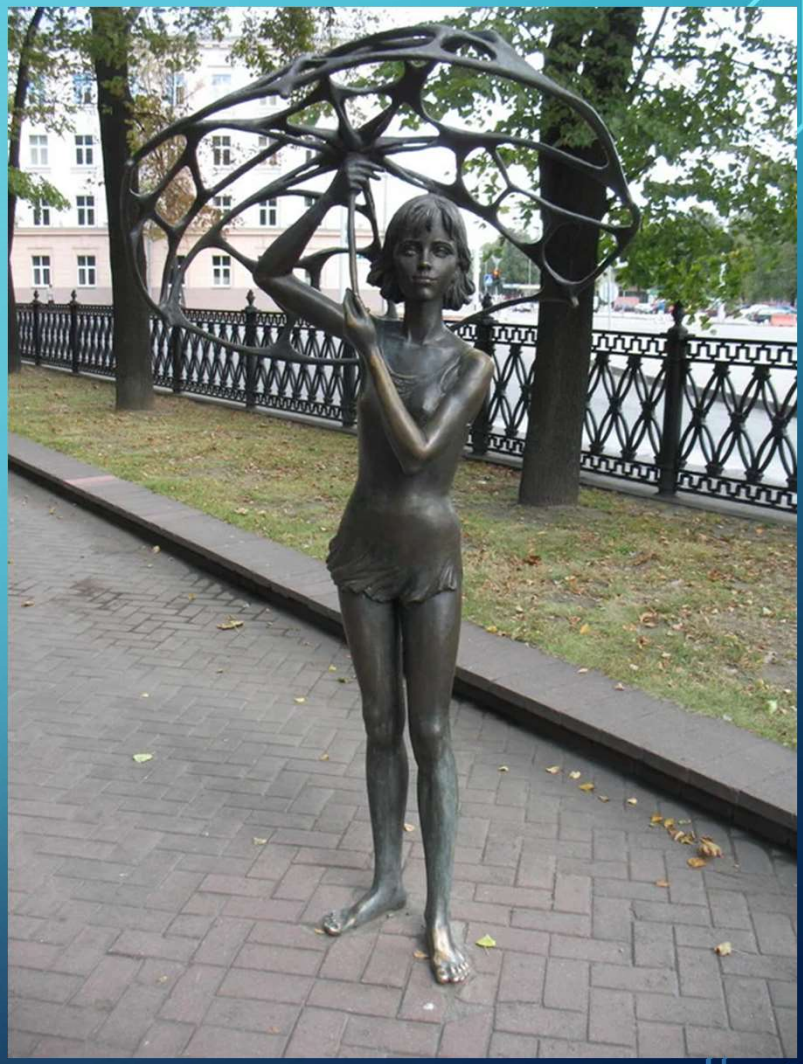


ТЕОРЕТИЧЕСКАЯ ИНФОРМАТИКА

- Дизайн признаков структурных объектов в задаче «структура-свойство» с использованием архитектур ***глубоких нейронных сетей***

Разделы теоретической информатики:

- *Анализ данных*
- *Большие данные*
- *Математические методы искусственного интеллекта, нейронных сетей.*



ГРАНТЫ РФФИ - 1

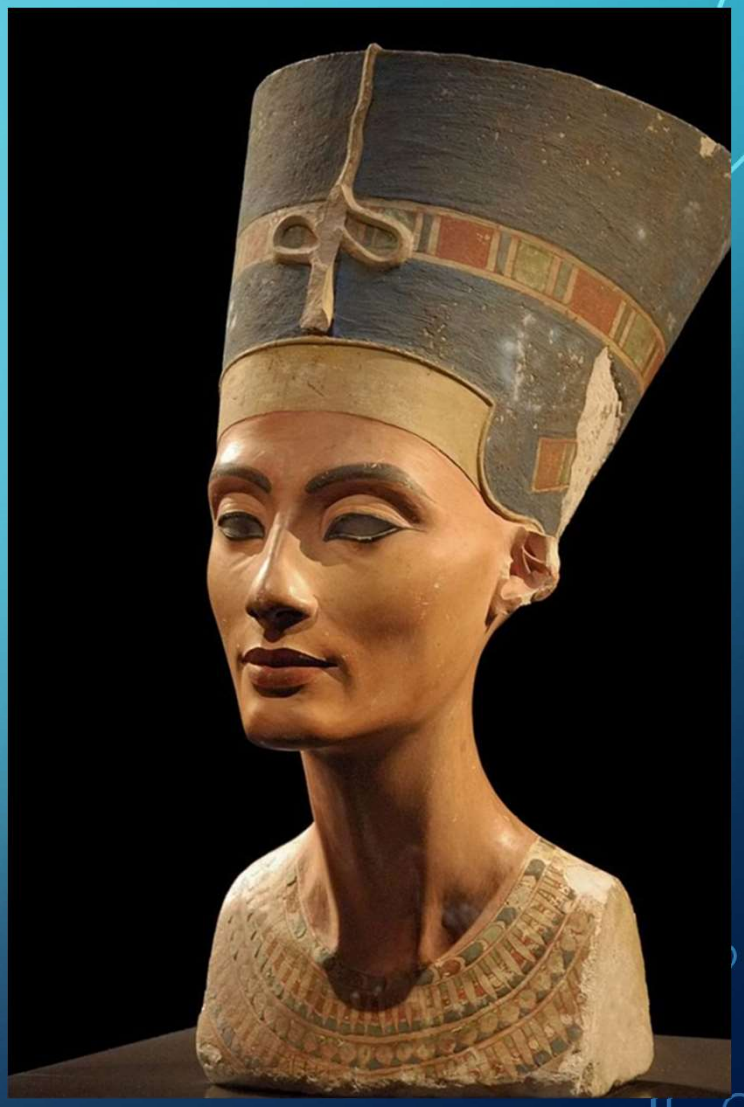
- РФФИ № 93-012-1045: "Унифицированные математические модели и программно-инструментальные системы для прогнозирования новых органических соединений с заданными свойствами";
- РФФИ № 94-01-00041: «Инструментальная система формирования баз знаний о зависимостях "структура-свойство" органических соединений на основе символьного представления фрагментов молекулярных графов»;
- РФФИ № 96-01-01598: «Распознавание пространственных форм молекул биологически активных соединений с целью компьютерного предсказания свойств новых веществ»;
- РФФИ № 97-07-90307: «Селекция метрик для поиска подобных молекул в структурных фактографических БД с использованием знаний “структура-свойство»;

ГРАНТЫ РФФИ - 2

- РФФИ № 98-01-00324: «Распознавание и классификация пространственных форм гибких молекул с использованием структурных символьных спектров и эволюционных алгоритмов»;
- РФФИ № 07-07-00282: «Система прогнозирования свойств химических соединений. Унифицированный репозиторий QSAR моделей и молекул»;
- РФФИ № 10-07-00694: «Развитие и реструктуризация репозитория моделей "структура-свойство" с целью проведения массового скрининга молекул»;
- *РФФИ № 19-07-00752: «Задача "Структура-свойство". Создание архитектур нейронных систем на основе прогнозирующих моделей обобщенного дерева решений»*

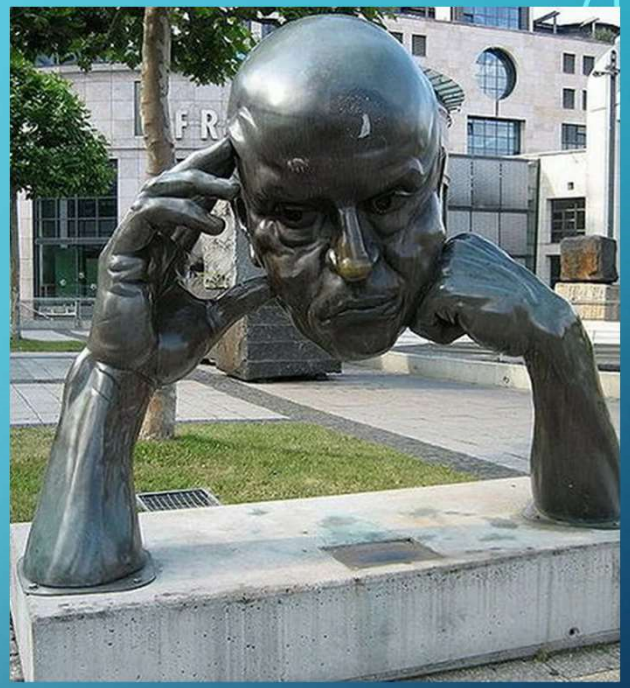
МАШИННОЕ ОБУЧЕНИЕ

- Программирование на ***Python*** — использование многочисленных библиотек машинного обучения
- Программирование нейронных сетей на ***TensorFlow и Keras***, — посмотрите в интернете!



ЗАДАЧА «СТРУКТУРА-СВОЙСТВО»

- Задано обучающее множество структурных объектов, для которых известно «целевое свойство» - имеем на входе множество пар (объект, свойство)
- Следует построить модель отображения объектов в «целевое свойство» с минимизацией ошибки на обучающем множестве.
- При появлении нового объекта по модели проводится прогнозирование «целевого свойства» или «***отказ от прогноза***».



КЛЮЧЕВАЯ ОСОБЕННОСТЬ СС-ЗАДАЧИ

- Заранее ***не задано*** пространство признаков исследуемых структурных объектов
- Это позволяет решать задачу не путем поиска новых алгоритмов, а проводить конструирование представления Объекта в модели, **адаптированное** под «целевое свойство» при заранее заданном методе прогнозирования

ОДР – ОБОБЩЕННЫЕ ДЕРЕВЬЯ РЕШЕНИЙ

- Для заранее заданного «прогнозатора» (например, для «***обобщенных деревьев решений***», созданных на ***кластерах*** обучающего набора данных) строятся семейства решений с ***индуктивным наращиванием сложности*** описаний объектов.
- Для поиска подмножества признаков, на которых получают «лучшие модели», используются такие ***эволюционные алгоритмы*** как Метод группового учета аргументов (МГУА)

МАРКЕРЫ ВЕРШИН ГРАФА

- Важным этапом конструирования признаков является выбор маркеров для ***символьной разметки вершин*** графов, которая меняет отношение ***эквивалентности фрагментов*** как инвариантов описания графов в задаче «структура-свойство»

МАРКЕРЫ ВЕРШИН ГРАФА - 2

- Маркеры вершин графов, описывающих структуру исследуемых объектов, кодируют ***локальные свойства объекта***.
- Для молекулярных графов маркеры могут кодировать интервалы таких ключевых физико-химических свойств как потенциал, липофильность, донорно-акцепторные факторы (способность отдавать или принимать электрон)

Вопросы?