

Темы КР и ВКР

Департамент анализа данных и искусственного интеллекта
2016-2017 учебный год

Список тем **курсовых работ** и **выпускных квалификационных работ**
бакалавров и магистров направления подготовки
"Прикладная математика и информатика"

Контактное лицо от департамента АДИИ:

Макаров Илья Андреевич

<http://hse.ru/staff/iamakarov>

Ilya Makarov

<http://hse.ru/en/staff/iamakarov>

К каждому преподавателю приложены 1-2 ссылки:

Ctrl + клик в оглавлении – переход на описание тем преподавателя

Ctrl + клик на ФИО в описании – переход на персональную страницу преподавателя (гиперссылка)

К некоторым темам приложены ссылки на доп. материалы.

Некоторые КР и ВКР ранжированы по разделам.

Дистанционные профессора иностранных университетов

Иностранные специалисты, работающие в НИУ ВШЭ

Руководители подразделений и ведущие сотрудники

Руководители большей части КР и ВКР департамента

Руководители специализированных тем или направлений

Руководители прикладных узкоспециализированных тем

Список тем КР и ВКР

Департамент анализа данных и искусственного интеллекта	1
Kertesz-Farkas Attila	3
Decrouez Geoffrey G.	6
Большакова Елена Игоревна	7
Вылиток Алексей Александрович	8
Жуков Леонид Евгеньевич	9
Захарьящев Михаил Викторович	10
Игнатов Дмитрий Игоревич	11
Ильвовский Дмитрий Алексеевич	14
Канович Макс Иосифович	15
Кузнецов Сергей Олегович	16
Макаров Илья Андреевич	17
Миркин Борис Григорьевич	23
Незнанов Алексей Андреевич	24
Объедков Сергей Александрович	27
Паринов Андрей Андреевич	28
Пьяных Олег Станиславович	29
Реброва Ольга Юрьевна	29
Строк Фёдор Владимирович	30
Черняк Екатерина Леонидовна	31
Яковлев Виктор Вадимович	32
Яворский Ростислав Эдуардович	33
Департамент психологии	34
Осадчий Алексей Евгеньевич	34
Vasily Klucharev/Boris Gutkin	34
Joseph W. MacInnes	34
Tadamasa Sawada	35
Zafer Iscan/Vadim Nikulin	35
Презентации проектов департамента психологии	36
Balaev Vladislav	36
Gutkin Boris , Lussange Johann	38
Joe MacInnes	41
Осадчий Алексей Евгеньевич	44
Sawada Tadamasa	47

Kertesz-Farkas Attila

Кертес-Фаркаш Атила

<http://www.hse.ru/en/staff/akerteszfarkas>

Contact: akerteszfarkas@hse.ru

Projects 1:

Human activity recognition from mobile phone sensors

[Распознавание деятельности человека на основе датчиков мобильных телефонов]

In this project student will implement a program for mobile phone (Android), which collects data from the sensors of the mobile phone (such as accelerometers) and then carries out some simple data analysis to recognize whether the user is running, walking, bicycling, or still. Student in this project will acquire hands-on experience with basic machine learning principles, data analysis methods, and with mobile phone developing tools.

Level: BSc

Project 2:

Bioinformatics: Learning generative probabilistic models for data identification of mass spectrometry data

[Биоинформатика: Обучение генеративных вероятностных моделей для идентификации данных масс-спектрометрии]

This project does not require a priori knowledge on biology.

The aim in this project is to learn to match a set of experimental data to their correct annotation. The matching will be implemented with Energy based models such as Restricted Boltzmann Machines, Bayesian networks. These models can then yield a probabilistic model which can be used to assess the probability of a correct matching. Algorithms and tests will be implemented and executed on Theano and GPU. Student in this project will acquire hands-on experience with generative models and basic deep learning techniques (Boltzmann machines, etc).

Level: BSc/MSc

Project 3:

Learning to generate music

[Генерация музыки при помощи машинного обучения]

In this project student is to develop generative programs which automatically composes some music. The generator process will be modelled with Recurrent Neural Networks, HMM, etc. and trained on some public music repositories. Algorithms and tests will be implemented and executed on Theano and GPU. In this project student will acquire skills with sequential data analysis and generative models. This project is in collaboration with NTRLab.

Level: MSc

Project 4:

Backpropagation methods for ReLU units in deep Artificial Neural Networks

[Обратное распространение ошибки для ReLU в глубоких нейронных сетях]

Backpropagation methods used to train ANN are general purpose methods which work with any kind of derivable activation functions. However, currently, ReLU units have received much attention. These are piecewise linear units. The aim of this project is to give a better formulation and code for backpropagation using ReLU. Algorithms and tests will be implemented and executed on Theano and GPU. In this project student will acquire skills on deep learning. This project is in collaboration with NTRLab.

Level: MSc

Project 5:**Mobile phone positioning****[Определение положения через мобильные устройства]**

Mobile phones connected to Base Transceiver Stations (BTS) and distance between phone and BTS can be estimated from the Strength of Received Signal (RSS). Therefore, when a mobile phone is connected to several BTSs its location can be determined. In this project student will learn basics of localization techniques. This project is in collaboration with Huawei Russian Research Center (in Moscow).

Level: BSc/MSc

Project 6:**Mobile phone tracking****[Отслеживание положения мобильного устройства]**

This project is similar to the previous one. The main difference is that, a series of measurements are given and changes in it can be used to improve the geo-localization of the mobile phone. In this project student will learn basics of localization techniques. This project is in collaboration with Huawei Russian Research Center (in Moscow).

Level: BSc/MSc

Project 7:**Learning to play computer games with OpenAI****[Машинное обучение в компьютерных играх на платформе OpenAI]**

OpenAI has released a general reinforcement learning package which enables to make a sequence of controls. The aim of this project is to try to create a program which could learn and play classic computer games such as pac-man, space intruders, etc. Algorithms and tests will be implemented and executed on Theano and GPU. In this project student will acquire skills with generative models.

<https://openai.com/blog/openai-gym-beta/>

Level: BSc/MSc

Project 8:**Learning to reconstruct 3D face****[Трёхмерная реконструкция лица методами машинного обучения]**

Learning to reconstruct 3D face from several two dimensional pictures. In this project several picture is taken of the head of a person from different angles, and the aim is to reconstruct it in 3D space. Algorithms and tests will be implemented and executed on Theano and GPU. In this project student will acquire skills with generative models.

Level: BSc

Project 9:**Learning generative models with side information****[Обучение генерирующих моделей с дополнительной информацией]**

Generative methods such as (Expectation-Maximization, Restricted Boltzmann Machines) are used to learn a subtle structure of a given data. In general, generative models do not take into account information about classes when it is available. This is often the case in deep learning applications. This project aims to add modification to generative models to take into account such side information. Algorithms and tests will be implemented and executed on Theano and GPU. In this project student will acquire skills with generative models.

Level: BSc/MSc

Project 10:**Pronunciation trainer****[Тренер произношения]**

Develop an (Android) app which helps users improve their pronunciations of English words. The core part of this project involves matching the sound produced by user to a correct pronunciation of the word. This problem can be tackled down with using methods such as Dynamic Time Wrapping, or using some Hidden Markov Models. In this project student will acquire skills with sequence alignment methods or generative models. Level: BSc/MSc

Project 11:

Speech recognition/synthesis

[Распознавание/синтез речи]

Student will develop a classic speech recognition and speech synthesis program. Student will gain experience with sequential data analysis, and hidden Markov models. Methods will be implemented on Theano, Python. In this project student will acquire skills with generative models.

Level: BSc/MSc

Project 12:

Learning String Kernels utilizing adaptive variable length compression (such as LZW)

[Обучение применению Строковых ядер Ядра с использованием адаптивного сжатия переменной длины (например, LZW)]

Student will develop a discrete kernel function for strings, where the atomic information will be extracted from LZW compressors. This method will be tested on protein sequence classification problems. In this project student will acquire skills with convolution kernels and kernels on discrete structures.

Level: MSc

Project 13:

Gesture recognition

[Распознавание жестов]

In this project student will implement an Android app which recognizes Latin letters written with a mobile phone in the air. Algorithms and tests will be implemented and executed on Theano and GPU. Student will acquire skills with sequential data analysis and generative models.

Level: MSc

Decrouez Geoffrey G.

Декруэ Жофри Жерар

<http://www.hse.ru/en/staff/gdecrouez>

Project 1.

The split-sample method

[Метод разделения выборки элементов]

The split-sample method was recently introduced as a way to reduce the amplitude of oscillations in the coverage probability of confidence intervals for the mean value of discrete distributions, including bootstrap intervals. In this project, the student will first learn about theoretical developments behind this method (normal approximation, Edgeworth expansions), then investigate numerically the performance of the split-sample method on a broader range of bootstrap intervals, and explore the possibilities to adapt the method to other settings, such as in hypothesis testing or with the jackknife. Interests in statistical theory and basic programming skills in Matlab or R are essential.

Level: 4th year/Master 1.

Project 2.

Analysis of scale-invariant time series

[Анализ временных рядов, инвариантных относительно шкалирования]

Time-series data presenting scale invariance do not possess a well-defined time scale. Instead, their dynamics are understood when studied across a whole range of scales. Examples of data with empirical scale-invariance include network traffic, financial timeseries, and other natural phenomena in physics and biology. The analysis of scale-invariant time series usually assumes that the local regularity of the process is homogeneous, that is, it does not change with time. We are interested here in situations where the local regularity may vary. Using the crossing tree, a tool providing a representation of the data adapted to its dynamics, we develop a strategy to detect such changes. The method will be compared with state-of-the-art techniques, and then applied on real data.

Level: 4th year/Master 1.

Большакова Елена Игоревна

Темы ВКР:

1. Методы извлечения информации о событиях из проблемно-ориентированных текстов
Methods of extracting information about the events from the problem-oriented texts
2. Методы извлечения терминологической информации из научно-технических текстов
Methods of extracting terminological information from scientific and technical texts
3. Исследование формальных критериев паронимии слов русского языка
Formal criteria of paronymia of Russian words

Темы курсовых работ:

1. Машинное обучение в задачах автоматической обработки текстов
Machine Learning for automatic text processing
2. Статистические модели естественного языка
Statistical models of natural language
3. Построение базы статистики синтаксических связей слов русского языка
Building a database of statistics syntax relations for Russian words
4. Автоматизированное пополнение базы словосочетаний
Automated reinforcement of phrases base
5. Усовершенствование модуля морфологического анализа слов русского языка
Improvement of the morphological analysis module for Russian words

Темы курсовых работ:

Статистические методы обработки текстов:

1. Частотный подход к решению задач автоматической обработки текстов
Frequency approach to solving problems of automatic text processing
2. Вероятностный подход к решению задач автоматической обработки текстов
Probabilistic approach to solving problems of automatic text processing
3. N-граммы и автоматическая обработка текстов
N-gramms and automatic text processing

Вылиток Алексей Александрович

1. Уменьшение циклической сложности конечного автомата
Reducing the complexity of a cyclic finite automaton
2. Алгоритмы сопряжения графовых описаний формальных языков
Algorithms of linking for graph descriptions of formal languages

Жуков Леонид Евгеньевич

Источник проектов 2011 года:

<http://leonidzhukov.ru/hse/2011/projects/>

Темы курсовых и дипломных работ

- Применение методов SNA [*Methods of Social Network Analysis*]
 - a. Анализ и визуализация финансовых сетей
Analysis and visualization of financial networks
 - b. Исследование взаимодействие сотрудников в организации
Research on the interaction of employees in the organization
 - c. Исследование сетей венчурного финансирование
Research on the networks of venture financing
 - d. Анализ и визуализация лингвистических сетей
Analysis and visualization of linguistic networks
 - e. Анализ и визуализация корпоративных сетей
Analysis and visualization of corporate networks
 - f. Исследование и визуализация human connectome
Research and visualization of human connectome
 - g. Исследование и визуализация family trees
Research and visualization of family trees
- Разработка алгоритмов анализа сетей
[*Development of algorithms for network analysis*]
 - a. Нахождение сообществ в социальных сетях
Finding communities in social networks
 - b. Моделирование распространения влияния
Simulation of influence spread
- Методы машинного обучения [*Machine Learning methods*]
 - a. Предсказательная аналитика
Predictive analytics
 - b. Рекомендательные системы
Recommended systems
 - c. Глубокое обучение
Deep learning
- Большие данные [*Big data*]
 - a. Алгоритмы семплирования
Sampling algorithms
 - b. Рандомизированные алгоритмы
Randomized algorithms
- Методы визуализации данных
Methods of data visualization
- Моделирование социальных процессов
Modeling social processes

Захарьящев Михаил Викторович

<http://www.dcs.bbk.ac.uk/~michael/>

1. Онтологические базы данных
Ontology-Based Data Access
2. Дескриптивная логика
Description Logic
3. Технологии семантической паутины
The Semantic Web Technologies

<http://cs.hse.ru/data/2015/05/20/1097298008/talk-HSE-seminar-topics.pdf>

Level: bachelor\graduate (4 year)\master\postgraduate

Игнатов Дмитрий Игоревич

Условные обозначения

Б – работа уровня бакалавриата

М – работа уровня магистратуры

П+ – требует реализации алгоритмов

П – требует программирования для предобработки данных

Задачи от компании ЗАО «Рекомендата» (Imhonet.ru)

1. Меры оценки качества рекомендательных систем. (Литературный обзор и эксперименты.) (БМП+) *Quality assement for Recommender Systems*
2. Рекомендательные системы на основе ассоциативных правил и матричной факторизации. (БПМ+) *Recommender Systems based on association rules and matrix factorisation*
3. Ансамблевые рекомендательные системы (БПМ+) *Similarity fusion in recommender systems*

Курсовые

1. Методики оценки качества рекомендательных систем. (БМ или БМП или БПМ+) *Methods of quality assesment in recommender systems*
2. Разработка лабораторной работы по рекомендательным системам. (БМП или БПМ+) *Development of a practical course in Recommender Systems*
3. Разработка лабораторной работы по анализу последовательностей. (БМП или БПМ+) *Development of a practical course in sequence mining*
4. Разработка лабораторной работы по применению многомерного шкалирования. (БМП или БПМ+) *Development of a practical course in Multidimensional Scaling.*
5. Исследование публикационной активности методами анализа формальных понятий и тематического моделирования. (БМП или БМП+) *Publication activity study by means of Formal Concept Analysis and Topic Modeling.*
6. Разработка лабораторной работы по анализу данных систем совместного доступа к ресурсам (Flickr, Bibsonomy, Delicious и др.). (БМП или БПМ+) *Development of a practical course in resource sharing systems (Flickr, Bibsonomy, Deliciois etc.)*
7. Разработка лабораторной работы по методам анализа социальных сетей. (БМ или БМП или БМП+) *Development of a practical course in Social Networks Mining.*
8. Разработка лабораторной работы по глубинному обучению. (БМ или БМП или БПМ+) *Development of a practical course in Deep Learning techniques.*
9. Решение реальных задач машинного обучения и разработки данных ресурса Kaggle. (БМП или БМП+) *Solution of real machine learning and data mining tasks from Kaggle*
10. Анализ данных в области образования (Educational Data Mining). (БМП или БМП+) *Educational Data Mining*

Темы ВКР

1. Исследование публикационной активности методами разработки данных (Data Mining). (БМП или БМП+)
Publication activity analysis by means of data mining.
2. Разработка данных в демографии и социологии. (БП или БП+)
Data mining and machine learning in social sciences
3. Разработка данных в образовании. (БП или БП+)
Educational data mining
4. Методы оценки качества рекомендательных систем. (БМП+)
Methods of quality evaluation for Recommender Systems
5. Контекстно-зависимые рекомендательные системы на основе трикластеризации (для анализа фолксономий). (БМП+)
Triclustering based context-aware recommender systems
6. Машинное обучение для анализа мнений пользователей Интернет-ресурсами. (БМП+)
Machine learning for opinion mining of web users
7. Консенсусная кластеризация на основе решеток понятий. (БМП+)
Consensus clustering based on Formal Concept Analysis
8. Би- и трикластеризация для анализа данных социальных сетей. (БП или БМП+)
Bi- and triclustering for Social Network Analysis
9. Рекомендательные системы на основе узорных структур и бикластеризации (БПМ+)
Recommender Systems based on Pattern Structures and Biclustering
10. Прогнозирование покупок в потоковых данных (БПМ+)
Purchase prediction in stream data
11. Онлайн-овые и распределенные алгоритмы мультимодальной кластеризации (MapReduce, ApacheSpark) (БПМ+)
Online and distributed multimodal clustering
12. Агрегация сходства в рекомендательных системах (БПМ+)
Similarity Fusion in Recommender Systems
13. Рекомендательные системы на основе доверия (БПМ+)
Trust-based Recommender Systems
14. Разработка данных и машинное обучение для демографических последовательностей (БМП+)
Demographic sequence mining and learning

Задачи компании «Имхонет»

1. Рекомендательные системы на основе ассоциативных правил и матричной факторизации *Recommender Systems based on association rules and matrix factorisation*

Данные.

Оценочные профили пользователей.

Что требуется компании.

Проверить разные модели и выбрать подходящие на данных о кино и книгах.

Гипотеза: стоит использовать алгоритмы, которые умеют выявлять редкие правила с большим уровнем достоверности (confidence).

2. Музыкальные рекомендации на основе моделей ранжирования на графах *Music recommendations based ranking model on graphs*

Данные.

Данные о музыке MSD, плейлисты пользователей социальной сети «ВКонтакте», данные lastfm скроблинга (см. термин <http://www.lastfm.ru/help/faq?category=99>), оценки imhonet.

Что требуется компании.

Рабочие модели ранжирования предпочтений пользователей, одновременно учитывающие большое множество очень разнородных данных.

В этой задаче с самого начала нужно иметь в виду проблему производительности.

<http://eagle.zju.edu.cn/~lijun/pdf/MM-2010-Bu.pdf>

3. Меры оценки качества рекомендательных систем. *Methods of quality assesment in recommender systems*

(Литературный обзор и эксперименты.)

Что требуется компании.

Понимание, в каких ситуациях какие метрики лучше использовать.

Интересно было бы просмотреть/проиндексировать максимум количества работ (или более-менее репрезентативную выборку, чтобы посмотреть временные зависимости - как менялись самые предпочитаемые метрики, и зависимость от контента.

Исследовать метод агрегирования оценок и его подстановки в факторный фильтр на базе онтологии книг (<https://dl.dropbox.com/u/49259983/p39-sieg.pdf>)

с данными Имхонета и сравнить результат с обычной фильтрацией (без метаданных).

Использовать готовый пакет для метода подстановки мета-данных непосредственно в коллаборативную схему (http://apex.sjtu.edu.cn/apex_wiki/svdfeature) через поиск коэффициентов линейной комбинации к группам факторов, определяемыми метаданными.

Подробнее (стр. 5 формула 1): <http://svdfeature.apexlab.org/wiki/images/d/d1/Svdfeature-manual.pdf> - опять-таки с данными Имхонета.

Сравнить подход с уже реализованными методами компаниями Имхонет.

Ильвовский Дмитрий Алексеевич

Темы курсовых и дипломных работ:

1. Автоматический анализ качества машинного перевода на основе синтаксических и дискурсивных характеристик текста и веб-майнинга
Automatic quality analysis of machine translation results based on syntactic and discourse text features and web mining
2. Анализ связности и "качества" текста с помощью инструментов дискурсивного разбора текста
Analysis of text coherence and quality based on discourse text features
3. Интеллектуальная кластеризация коротких текстов: обзор и сравнение методов
Intellectual text clustering of short texts: survey and comparison of existing methods

Канович Макс Иосифович

<https://www.eecs.qmul.ac.uk/~mik/>

1. Формальные методы в классическом искусственном интеллекте для задачи планирования, но с количественными ограничениями времени
Formal methods in the traditional AI planning but with quantitative time constraints
2. Формальные методы верификации программного обеспечения
Formal verification of software. Effective logical formalisms for resource- and memory-sensitive reasoning
3. Формальные системы защищенного обмена информацией в рамках совместных систем и протоколов безопасности с количественными ограничениями времени
Formal systems for assured information sharing within collaborative systems and security protocols with quantitative time constraints
4. Формальные системы в компьютерной лингвистике
Formal systems in computational linguistics
5. Проверка логических моделей
Model Checking

<http://cs.hse.ru/data/2015/05/20/1097298008/talk-HSE-seminar-topics.pdf>

Level: bachelor\graduate (4 year)\master\postgraduate

Кузнецов Сергей Олегович

Примерные темы курсовых работ и ВКР бакалавров и магистров

1. Методы машинного обучения и майнинга данных в анализе медицинской информации
Methods of machine learning and data mining in medical informatics
2. Построение таксономий предметных областей по коллекциям документов
Constructing taxonomies of subject domains by collections of documents
3. Развитие алгоритмов анализа данных в программной системе FCART
Designing algorithms of data analysis in FCART software system
4. Анализ сходства пациентов по медицинской документации
Analysis of patients similarity by medical documentation
5. Быстрые методы кластеризации данных на основе решеток понятий
Fast methods of biclustering based on concept lattices
6. Быстрые алгоритмы анализа формальных понятий
Fast algorithms of Formal Concept Analysis
7. Методы машинного обучения и майнинга данных на основе решеток формальных понятий
Methods of machine learning and data mining based on concept lattices
8. Методы и алгоритмы анализа протоколов стратегических компьютерных игр
Methods and algorithms of protocol analysis for strategic computer games
9. Быстрые алгоритмы бикластеризации
Fast Algorithms for biclustering
10. Обучение архитектуре нейронной сети для глубокого обучения
Structures of Neural Networks for Deep Learning

Новые темы 2016:

1. Пилотный проект по применению машинного обучения в задачах прогноза потерь по клиентам банка
Pilot project for loss given default estimation and machine learning
2. Построение альтернативных предиктивных моделей дефолта по заемщикам
Alternative models for probability of default prediction in banking
3. Использование данных социальных сетей в задаче кредитного скоринга
Credit scoring based on social network data
4. Grid search по параметрам алгоритма "lazy classification" – параллельные вычисления
Grid search, parameter tuning and parallelization in "lazy classification"
5. Grid search по параметрам алгоритма "lazy regression" - параллельные вычисления
Grid search, parameter tuning and parallelization in "lazy regression"

Макаров Илья Андреевич

Проекты 2016-2017

1. Проектирование трехмерного шутера от первого лица для виртуальной реальности
Developing FPS Game for VR
(Unreal Engine 4, Strong English, C++ Coding. Learn AI Books)
2. Реконструкция трехмерной сцены по фото/видео
3D-Scene Reconstruction using photo/video
(Computer Vision, Strong English, C++ Coding. To be ready to learn 3D-Max and/or UE4)
3. Дополненная реальность на основе данных в Web-камеры
Augmented Reality based on Web-cam data
(Computer Vision, Strong English, C++ Coding. To be ready to learn 3D-Max and/or UE4)
4. Программный комплекс для хранения, категоризации и обработки мульти-медиа коллекций
Software system for storing, categorizing and processing multimedia collections
(Description Logic, XML, SQL, C#, Site Creating, Strong English)

Каждая из тем может быть темой курсовой работы, ВКР и магистерской диссертации, а в идеале – является сквозной темой научной работы студента на всех этих этапах.

**Язык выполнения: русский/английский,
вне зависимости от языка описания темы.**

Общие темы:

1. Программирование игрового искусственного интеллекта на движке UE4

Programming game artificial intelligence in UE4

2. Программирование игр на движке UE4, Blender. Создание модификаций с улучшением алгоритмов ИИ

Programming games in UE4, Blender. Making modifications to improve AI algorithms

3. Программирование поведения ИИ при распространении звука в лабиринте

Programming AI for the sound wave propagation in a maze

Построение модели, улучшение алгоритмов для распространения различных звуков в различных средах и окружениях с целью достижения реалистичного поведения ИИ; программирование графического интерфейса.

4. Программирование игровых минимаксных задач

Programming minimax games problems

Построение модели, создание, модифицирование и улучшение алгоритмов для поставленной задачи с целью достижения максимальной реалистичности игрового процесса; программирование графического интерфейса.

5. Анализ графа публикаций научно-педагогических работников НИУ ВШЭ

The publication graph analysis of scientific and pedagogical NRU HSE staff

6. Анализ распространения информации в социальных сетях

Information spread in Social Networks

7. Свободные темы

Теоретические темы

1. Computational Logic (Course projects in English)

- a) Logic and **Music** (Ingolf Max, The Logic of Chords and Harmony). Music experience required.
- b) Logic and **Quantum Theory** (Bob Coecke, Natural language meaning). Knowledge in Quantum mechanics basics will be plus for candidates.
- c) Logic and **Color** (Dany Jaspers, Colour lattices). The work will be devoted to segmentation of images and analyzing harmony of art in terms of colours' properties. A theory of ordered sets and lattices is required.
- d) Logic and **Nonsense** (Thomas Ferguson, Meaningless expressions). NLP knowledge is a plus for candidate.
- ~~e) Logic and **Fiction** (John Woods, Fiction Semantics). Descriptive works. Large experience in fiction books is required.~~
- ~~f) Logic and **Politics**. Modal logics.~~
- g) Dualization of a monotone Boolean function represented by a conjunctive normal form (CNF) (Kazuhisa Makino).

2. Задачи дискретной математики

- a) Открытые проблемы многозначной логики
Open problems of many-valued logic

(потребуется углубленное изучение булевой и многозначной логики с помощью научного руководителя)

- б) Алгоритмы и задачи на графах
Algorithms on graphs

(результаты могут быть использованы в дальнейших научных исследованиях у других научных руководителей при практической реализации, например различные варианты поисков A^* , B^*)

- в) Изучение свойств алгоритма или структуры данных, проблемы сложности алгоритмов и их реализаций
Research of the properties of the algorithm or data structure and its complexity
- (пример: инкрементное построение декартового дерева по гипотезе о распределении данных)

3. Задачи теории чисел

- a) Открытые проблемы многомерных обобщений цепных дробей
Open problems of multidimensional generalizations of continued fractions
- б) Аналитическая теория чисел. Получение асимптотик работы известных алгоритмов
Analytical theory of numbers. Obtaining asymptotics of known algorithms

4. Свободные темы по разделам высшей математики

Проектные темы КР и ВКР:

1. Построение модели игрового искусственного интеллекта на основе деревьев поведения, и обучающейся системы принятия решений, основанной на правилах, в Unreal Engine 4.

Апробация проекта проходит на примере участия написанного бота в игровом тесте Тьюринга, определяющего различимость игрока под управлением человека и компьютера.

Научная апробация должна быть представлена в виде работы на международной конференции / конференции по разработке компьютерных игр.

Общие требования: умение проектировать сложные системы данных и графически их представлять, навыки визуального программирования блюпринтов, личная заинтересованность и хорошие навыки самообучения, владение английским языком на уровне intermediate. Проект подразумевает тесное сотрудничество с научным руководителем. Занятость в проекте требует много времени; в случае академической неуспеваемости возможно оказание консультаций по математическим предметам с целью компенсировать пробелы.

Проект состоит из нескольких направлений, требующих каждая соответствующих компетенций.

1. Написание модуля визуального распознавания на C++.
2. Реализация обучения выбору путей и сглаживание путей на основе данных передвижений реальных игроков.
3. Адаптация стрельбы под человеческие рефлексy.
4. Сравнение eye-tracking и mouse-tracking при идентификации врага / объектов.
5. Выбор оружия и бонусов на основе reinforcement learning.
6. Логическое сравнение деревьев поведения на основе системы правил.

Работа осуществляется в рамках подачи итоговой модели БОТа на международный грант от компании Epic Games, а также индивидуального исследовательского проекта НИУ ВШЭ. Возможно создание проектно-учебной группы по результатам участия в конкурсе с оплатой результатов работы в случае успеха.

Description in English:

Logical Modelling of Behavior Trees for a Human-Like Bot in a First Person Shooter Game

Many computer games have computer-controlled agents playing against human players. The behavior of these non-playable characters (NPC) is described in terms of the artificial intelligence (AI) in the game. The AI is an important part of the game that provides important features of augmented virtual reality, such as human-like behavior of NPCs and interaction with them as if they were real. Entertainment industry provides a vast variety of different games, every time making an attempt to make a virtual world look more real. In modern games developers pay more attention to graphics quality and care less about the non-scripted behavior of NPC, which is an enemy to a human player (so called BOT). BOTs do not commonly use visual recognition during gameplay; they know in advance the structure of a map and the goal location. This concept leads to the fact that such computer players have an advantage over users. Multiple algorithms are presented to compensate BOT supremacy over human player, such as simulated time delay to fire backwards, bad accuracy during shooting process, almost infinite health level and endurance of a human player and many other methods to retain balance between an entertainment and a game challenge in virtual world.

In this project, we continue development of visual recognition model for the BOT due to which the computer player will be more realistic. We have already created a BOT, which operates in 3D-maze and uses its knowledge of the maze structure (locations of walls and obstacles). The AI detects dangerous zones, in which enemy presence is more likely. Also, it will use statistical parameters, which will define its behavior: smooth movement between control points, angle of view, physical movement parameters and other aspects. In order to achieve diverse BOT difficulty levels we construct a specific model with the following objects and models:

- Confidence intervals for time delay and sighting accuracy using visual recognition;
- Target sighting on fire recognition with adjusted parameters;
- Statistical model for the curves of the first-time and the second-time (correction) targeting accuracy and delays;
- Adaptation of shooting process with respect to recoil;

We will describe rule-based decision making system for priority actions with respect to the map of game events and implement it as root system for behavior trees, which controls BOT actions. We will try to compare such obtained behavior trees using interpretation of rule-based system as closed classes of many-valued logic, in particular, projection logic that allow us to analyze decision making models by true-false output schemata.

2. Построение графического интерфейса для отображения независимого экспертного мнения на базе данных фактов с учетом конфликтов интересов, принадлежности различным группам субъектов высказываний с возможностью автоматической обработки текста.

Апробация проекта проходит на примере информационного освещения Чернобыльской катастрофы 26 апреля 1986.

Общие требования: умение проектировать сложные системы данных и графически их представлять, личная заинтересованность на реализацию релизной версии проекта. Хорошее знание отечественной истории, владение английским языком приветствуются. Проект подразумевает тесное сотрудничество с научным руководителем. Занятость в проекте требует много времени; в случае академической неуспеваемости возможно оказание консультаций по математическим предметам с целью компенсировать пробелы.

Проект состоит из нескольких направлений, требующих каждая соответствующих компетенций.

1. Проектирование реляционной базы данных фактов событий и действий людей с учетом вероятностного характера истинности высказывания в зависимости от принадлежности субъекта высказывания различным группам (возраст, профессия, личная или корпоративная заинтересованность, др.)

Требования: умение программировать распределенные абстрактные типы данных с возможностью модификации структуры хранения данных в процессе развития проекта; написание кода, независимого от ОС

2. Проектирование графического интерфейса для отображения независимого экспертного мнения по базе данных фактов.

Требования: умение программировать графический интерфейс высокого уровня сложности с возможностью его усовершенствования в процессе развития проекта. Предполагается использование WinAPI.

3. Оцифровка видео-/аудио- материалов путем составления текстов разговоров и субтитров, заполнение

базы данных фактов на основе составления краткого содержания текстовых материалов (книги, статьи, субтитры к видео-/аудио- материалам). Перевод всех материалов на английский язык.

Требования: умение обрабатывать большие объемы знаний, высокая скорость чтения, хороший английский (возможно владение переводчиками с ручно правкой пос лова), хорошее знание истории.

4. Применение известных алгоритмов автоматической обработки текста с целью сравнения с независимой экспертной оценки, полученной в ходе работы над проектом.

Требования: умение быстро обучаться новым программным продуктам и адаптировать их, синтезируя с методами программирования, используемыми в проекте.

Пункты 1.-2. неразрывно связаны и требуют первичной реализации. После этого пункты 3. и 4. являются обособленными самостоятельными заданиями в рамках проекта и могут быть выбраны отдельными студентами независимо.

3. Построение графического интерфейса для отображения независимого экспертного мнения относительно экземпляров конкретных типов данных, и установления связей между ними с возможностью автоматической обработки.

Апробация проекта проходит на примере классических типов данных для хранения информации: тексты, видео, фото, аудио, музыка, программы, игры.

Общие требования: умение проектировать сложные системы данных и графически их представлять, личная заинтересованность на реализацию релизной версии проекта. Владение английским языком приветствуются. Проект подразумевает тесное сотрудничество с научным руководителем. Занятость в проекте требует много времени; в случае академической неуспеваемости возможно оказание консультаций по математическим предметам с целью компенсировать пробелы.

Проект состоит из нескольких направлений, которые после выбора типа данных, требуют:

1. Проектирование реляционной базы данных конкретного типа;
2. Проектирование графического интерфейса для отображения все полноты информации для данного типа данных и отображения связей и зависимостей конкретного экземпляра данных от соответствующих параметров
3. Оцифровка видео-/фото-/аудио- материалов путем составления текстов разговоров и субтитров, заполнение базы данных на основе составления краткого содержания текстовых материалов (книги, статьи, субтитры к видео-/аудио- материалам). Перевод материалов на английский язык желателен.
4. Применение известных алгоритмов автоматической обработки текста с целью выяснения наличия связей между описаниями различных экземпляров данных

После выбора типа данных требуется реализовать все пункты данной программы. Использование общих наработок для базы данных и абстрактного графического интерфейса возможно для выполнения различными студентами заданий по разным типам данных.

5. Свободные темы по инициативным проектам студентов.

Проектные темы для 1 курса:

<http://www.cs.hse.ru/ai/stud-projects>

[Проектирование 2D-игр с использованием MFC](#)

[Тестирование тактики кооперативной игры в трехмерном лабиринте на Unreal Engine 4](#)

Задания на практику:

[Анализ графа публикаций научных сотрудников НИУ ВШЭ](#)

[Визуализация выпуклых оболочек целых точек в полиэдрах](#)

[Изучение диаграммы Поста и расширенной теоремы Поста](#)

[Изучение полных систем тождеств и теоремы Линдона](#)

[Изучить пакеты BibTeX, Multibib, Bibunits, Hyperref, оболочку Jabref](#)

[Моделирование 3D шутера Counter Strike 1.6](#)

[Описание паттернов поведения игрока в трехмерном шутере](#)

[Основы программирования на движке Unreal Engine 4 – 1 \(передвижение и стрельба\)](#)

[Основы программирования на движке Unreal Engine 4 – 2 \(выбор оружия и обучение\)](#)

[Подготовка презентаций по онлайн курсу, посвященному анализу социальных сетей](#)

[Проверка корректности заданий](#)

Миркин Борис Григорьевич

1. Построение и анализ графов референций на ключевых словах для коллекций неструктурированных текстов
Designing and analyzing reference graphs over keywords for collections of unstructured texts
2. Выявление характера цитирования в научных статьях по материалам Соционет.ру
Analysis of the citation type in research papers using Socionet.ru
3. Изучение, сравнение и разработка методов, инспирированных природой, для кластер-анализа
Learning, comparing and designing nature-inspired methods for clustering
4. Программирование интеллектуального метода к-средних на языке Питон
Coding the method of intelligent k-means clustering
5. Методы кластер-анализа и выявления сообществ на сетях и их сравнение
Methods for clustering and community detection in networks

Незнанов Алексей Андреевич

Курсовые

1. Алгоритмы и программные средства анализа неструктурированных коллекций данных под управлением онтологии.
 - a. Algorithms and Software Tools for Ontology-controlled Analysing of Unstructured Data Collections
2. Хранилища неструктурированных данных и их использование: форматы данных, запросы, оптимизация.
 - a. Unstructured Data Storages and Applications: Data Formats, Queries, Optimization
3. "Облачные" технологии и их использование для масштабирования алгоритмов интеллектуального анализа данных.
 - a. Cloud technologies and their Application for Intelligent Data Analysis Algorithms Scaling.
4. Браузер внешних источников данных для промежуточного хранилища аналитической системы.
 - a. External Data Sources Browser for Intermediate Storage of Analytical Software.
5. Создание интерактивных алгоритмов решения трудно решаемых задач структурного анализа.
 - a. Development of Interactive Algorithms of Computationally Hard Problems of Structural Analysis.
6. Анализ сходства графовых моделей систем на основе моделей структурной сложности.
 - a. Graph Models Similarity Analysis Based on Mathematical Models of Structural Complexity.
7. Создание библиотек алгоритмов порождения и анализа комбинаторных конфигураций на структурах.
 - a. Development of Combinatorial Configuration Construction Algorithms Implementations Libraries in Applied Graph Theory.
8. Разработка специализированных версий алгоритмов для анализа социальных сетей.
 - a. Development of Specialized Version of Structural Analysis Algorithms for Social Network Analysis.
9. Реализация и исследование современных алгоритмов химической информатики.
 - a. Implementation and Comparative Analysis of Modern Cheminformatics Algorithms.
10. Визуализация структурной информации: методы и алгоритмы.
 - a. Structural Analysis Visualization: Methods and Algorithms.
11. Свободно-распространяемые АСНИ в области дискретной (конструктивной) математики.
 - a. Free Research Software Tools in the Field of Discrete Mathematics.
12. Интеграция систем управления академической библиографией в *Microsoft Word*.
 - a. Integration of Academic Bibliography Management Systems in Microsoft Word.
13. Инструменты управления стилями документа в формате *Open Document Format*.
 - a. Development of Software Tools for Document Style Management in Open Document Format.
14. Симулятор корпусов НИУ ВШЭ (3D).
 - a. Development of NIU HSE Buildings 3D-simulator.

ВКР

1. Алгоритмы и программные средства анализа коллекций неструктурированных данных под управлением онтологий (в том числе на базе АЧИ «*Formal Concept Analysis Research Toolbox*» – FCART).
 - a. Индексация текстов и полнотекстовый поиск: эффективная реализация и интеграция с подсистемой запросов системы FCART.
 - i. Text Indexing and Full Text Search: Effective Implementation and Integration in IDS Query Subsystem of FCART.
 - b. Специализированные языки запросов к метаонтологиям и предметным онтологиям для системы FCART.
 - i. Specialized Metaontology and Query Languages of FCART.
 - c. Эффективная реализация и сравнительный анализ алгоритмов интеллектуального анализа данных в объектно-признаковом представлении для системы FCART.
 - i. Effective Implementation and Comparative Analysis of Intelligent Data Analysis Algorithms for Object-Attribute Representation of Datasets.
 - d. Динамический анализ данных в объектно-признаковом представлении: проблемы и решения.
 - i. Dynamical Data Analysis for Object-Attribute Representation of Datasets.
 - e. Развитие подсистемы поддержки скриптов на интерпретируемых языках (*DWScript*, *Python*) для системы FCART.
 - i. Advancing of scripting subsystem of FCART.
 - f. Подсистема генерации синтетических аналитических артефактов для системы FCART.
 - i. Development of Synthetic Artefacts Generation Subsystem of FCART.
2. Задачи анализа данных в области медицинской информатики.
 - a. Интегрированная среда поддержки процессов контролируемого клинического эксперимента.
 - i. Development of Integrated Software Environment to Support Conducting of Randomized Clinical Trial.
3. Имитационное моделирование систем управления "умным домом" (*smart house*).
 - a. Построение и тестирование модели пользователя (жильца) "умного дома".
 - i. Development and simulation of Smart House User Model.
 - b. Построение и тестирование модели определённого устройства "умного дома".
 - i. Development and simulation of Smart House Device Model.
 - c. Разработка алгоритма управления компонентами "умного дома" и его имитационное моделирование.
 - i. Development and simulation of Smart House Controller Model.
4. Оптимизация графа сценариев использования при проектировании человеко-машинных интерфейсов (по областям).
 - a. Use Case Graph Optimization in Human-Computer Interaction Design.
5. Методы и алгоритмы эффективного решения базовых задач структурного анализа: различения, анализа симметрии и сходства, задач о покрытиях и др.
 - a. Точное и приближённое решение задач различения графов: алгоритмы и их вычислительная сложность.
 - i. Precise and Approximate Solutions for Graph Discrimination Problem: Algorithms and Computational Complexity.
 - b. Исследование различных типов порождающих множеств групп автоморфизмов графов и вычислительная сложность использующих их алгоритмов.
 - i. Algorithms and Computational Complexity for Various Kinds of Graph Automorphism Group Generative Sets.

- c. Задачи о покрытиях: классификация, подходы к решению, алгоритмы.
 - i. Graph Covering Problems: Classification, Approaches, Algorithms.
- 6. Визуализация структурной информации: методы и алгоритмы.
 - a. Алгоритмы (включая интерактивные) эффективной автоматической прорисовки решёток (линейных диаграмм) и их фрагментов.
 - i. Effective Algorithms for Automatic Line Graph and Their Fragments Diagram Construction.
 - b. Методы визуализации семантических сетей и их применение.
 - i. Visualization of Semantic Networks.
 - c. Ортогональная прорисовка обыкновенных графов с вершинами, степень которых больше 4.
 - i. Orthogonal Graph Drawing for Vertex Degree > 4 .
 - d. Адаптивные алгоритмы прорисовки различных классов графов методом физических аналогий.
 - i. Adaptive Graph Drawing Algorithms based on Physical Models.
 - e. Алгоритмы прорисовки графов с учётом симметрии расположения фрагментов.
 - i. Graph Drawing Algorithms Taking into Account Fragments Layout Symmetry.
 - f. Интерактивная прорисовка транзитивных графов.
 - i. Interactive Algorithms for Transitive Graph Drawing.

Список тем, не разделенный на категории (КР и ВКР)

1. Алгоритмы построения базисов импликаций

Algorithms for computing implication bases

Импликация здесь — это выражение вида $A \rightarrow B$, где A и B — подмножества признаков, интерпретируемое следующим образом: "Всякий объект, обладающий всеми признаками из A , обладает также всеми признаками из B ". Базис импликаций для массива объектно-признаковых данных — это подмножество импликаций, справедливых для объектов этого массива, обладающее некоторыми свойствами (полнота и избыточность).

2. Алгоритмы машинного обучения с запросами

Algorithms for learning with queries

Алгоритмы машинного обучения с запросами. Такие алгоритмы не просто используют фиксированный набор данных для обучения, но могут в процессе обучения задавать вопросы и получать на них ответы. Хороший алгоритм должен понимать, какие вопросы нужно задавать в каждый конкретный момент, чтобы процесс обучения был максимально эффективным.

3. Алгоритмы обучения предпочтениям

Algorithms for preference learning

Задача таких алгоритмов состоит в том, чтобы по имеющимся данным о конкретных объектах и предпочтениям на них построить модель, позволяющую предсказать предпочтения на новых объектах.

4. Исследование семантических полей при помощи методов формального анализа понятий

Exploration of semantic fields using formal concept analysis

Целью работы является создание методологии (и ее программной реализации), которая могла бы использоваться лингвистами в процессе исследования структуры семантических полей на материале различных языков. Предполагается использовать методы, позволяющие построить по лингвистическим данным так называемую решетку формальных понятий — своеобразную карту исследуемого семантического поля, из которой видно, каким образом различные значения, свойственные лексемам этого поля, группируются в различных языках.

5. Методы фильтрации шума в решетках формальных понятий

Methods for noise reduction in concept lattices

Формальный анализ понятий — раздел теории решеток с приложениями в области анализа данных. Одной из основных проблем применения методов формального анализа понятий на практике является их слабая устойчивость к шуму в данных. Поэтому использование решетки понятий, построенной по реальным данным, в качестве таксономии некоторой предметной области подразумевает решение задачи идентификации и удаления узлов решетки, возникших в результате шума.

Паринов Андрей Андреевич

Темы для КР:

1. Разработка плагинов для Trac
The development of plug-ins for Trac
2. Исследование языка Scala
Research of Scala language
3. Исследование ontology to access data systems
Research of ontology to access data systems
4. Исследование api электронных библиотек
Research of digital libraries api
5. Цифровая подпись и ldap системы
Digital signature and ldap system
6. Разработка системы сбора данных twitter, facebook, linkedin
Developing a system of data collection from twitter, facebook, linkedin
7. Исследование системы Apache Spark
Reserach of Apache Spark
8. Исследование системы Hadoop
Reserach of Hadoop
9. Исследование MS Sharepoint
Reserach of MS Sharepoint

Пьяных Олег Станиславович

Прикладная область - Медицинская информатика.

<http://ami.hse.ru/en/news/96056728.html>

<http://www.droleg.com/MI>

Примерные темы (направления исследований):

1. Алгоритмы сжатия данных с потерями (lossy), с контролируемой погрешностью.
Применение для диагностического сжатия медицинских изображений
Data compression algorithms with losses (lossy), with controlled accuracy. Application for diagnostic medical image compression
2. Моделирование и выделение поверхностей в 3-мерных объемах, в реальном времени.
Применение для распознавания и измерения опухолей.
Modeling and recovering surfaces in three-dimensional volumes in real time. Application to the detection and measurement of the tumors
3. Алгоритмы нелинейного удаления шума из 2-х и 3-х мерных изображений. Повышение диагностического качества медицинских данных
Algorithms for nonlinear noise reduction of 2-D and 3-D images. Improving the quality of medical diagnostic data
4. Улучшение качества компьютерных томограмм (изображений полученных преобразованием Радона)
Improving the quality of CT images (images obtained by the Radon transform)

Сложно:

насколько возможно восстановить 3-мерное изображение из 2-мерного рентгеновского снимка?

5. Способы восстановления 3-мерного изображения из 2-мерного рентгеновского снимка
Methods of restoring 3-dimensional images from two-dimensional X-ray

Олег Станиславович - преподаватель Медицинской школы при Гарвардском Университете и постоянный участник рабочей группы Международного Комитета DICOM (Индустриальный Стандарт создания, хранения, передачи и визуализации медицинских изображений и документов обследованных пациентов).

<http://cs.hse.ru/data/2015/05/20/1097298721/HSE.MI.pdf>

Реброва Ольга Юрьевна

Темы КР и ВКР:

Разработка компьютерных систем медицинской диагностики и прогнозирования с применением методов майнинга данных

Development of computer systems for medical diagnosis and prognosis using data mining

Строк Фёдор Владимирович

Все темы могут иметь сложность от курсовых работ (со 2 курса бакалавриата) и до магистерских диссертаций.

1. Написание утилиты для сравнения содержимого баз данных
Programming utility for comparison of database contents
2. Data mining и БД: разработка и реализация методов анализа данных из БД
Data mining and database: development and implementation of methods for analyzing the data from the database
3. Сравнительный анализ реляционной и объектно-ориентированной моделей данных
Comparative analysis of relational and object-oriented data models

Литература:

1. CrowDM: the System for Collaborative Platform Data Analysis
D.I. Ignatov, A.Y. Kaminskaya, A.A. Bezzubtseva, E.L. Chernyak, K.N. Blinkin, D.R. Nedumov, O.N. Chugunova, A.V. Konstantinov, N.S. Romashkin, F.V. Strok, D.Goncharova, R.E. Yavorsky
В кн.: EEML 2012 – Experimental Economics in Machine Learning, 2012. С. 61—71
2. Comparing and analyzing the computational complexity of FCA algorithms A.A. Neznanov, F.V. Strok
// Proceedings of SAICSIT 2010 - Annual Research Conference of the South African Institute of Computer Scientist and Information Technologists, South Africa: ACM Press, 2010.
3. Comparing and analyzing the computational complexity of FCA algorithms A.A. Neznanov, F.V. Strok
// Proceedings of Annual Research Conference of the South African Institute of Computer Scientist and Information Technologists (SAICSIT 2010), 2010. С. 417—420
4. Consensus issues in dynamics F.V. Strok 2010.
5. Оценка и сравнение производительности алгоритмов анализа формальных понятий Строк Ф.В., Незнанов А.А. // Двенадцатая национальная конференция по искусственному интеллекту с международным участием. КИИ-2010. Труды конференции., Москва: Физматлит, 2010.

Черняк Екатерина Леонидовна

Общие направления КР и ВКР:

1. Извлечение временных меток из русскоязычных текстов
Timestamps extraction from Russian texts
Python + NLTK
2. Изменение мнения о продукте со временем
Extraction of time changing opinion
Python + NLTK, scraping
3. Тренды в публикациях по CS, NLP
Trend detection in CS or NLP research papers
Python + NLTK
4. Извлечение метафор из нехудожественных текстов
Extraction of metafora from non fiction
Python + NLTK
5. Сравнение методов иерархической классификации на данных YaCa
Comparison of hierarchical classification methods on YaCa data
Python + NLTK + scipy
6. Использование строковых структур в задачах обработки и анализа текстов
Using string structure for text processing and text mining
желание и возможность много программировать

Для выполнения работ в области автоматической обработки текстов требуется знание языка Python (или готовность его изучить) и готовность работать с англоязычными источниками.

Темы КР и ВКР 2014/15:

1. Сбор, обработка, анализ и визуализация веб-данных (новостных потоков, Twitter, Википедии, блогов);
2. Методы выявления скрытых тем (latent topics): обзор и эксперименты;
3. Методы построения таксономий (в т.ч. по данным конкурса SemEval) и выявления таксономических отношений в текстах;
4. Разделение сложных слов в немецком языке на составляющие;
5. Обработка и анализ данных клинических испытаний (совместно с сотрудниками лаборатории клинической психологии и психотерапии Московского НИИ Психиатрии).
6. Проведение прикладных исследований на платформе SAS (SAS Analytics).

Яковлев Виктор Вадимович

Темы курсовых работ:

1. Тренажер по русскому языку, основанный на выравнивании текстов
Russian Language Trainer based on Strings Alignment

Предлагается применить подход, традиционно используемый в биоинформатике для глобального выравнивания аминокислотных последовательностей, к задаче сравнения текстов на естественном (русском) языке. Конечная цель - веб-приложение, предназначенное для автоматизированной проверки диктантов и визуализации результатов. Особенностью этой задачи является то, что набранный пользователем текст может значительно отличаться от эталонного, и таким образом нужно уметь находить в нем (за приемлемое время работы) как замены отдельных символов или слов, так и пропуски, либо некорректно вставленные цепочки символов. Работа является гибридом как научно-исследовательской деятельности (нужно исследовать применимость методов и подобрать необходимые параметры), так и производственного проекта (нужно получить законченный программный продукт).

1. Владение базовыми алгоритмами и методами динамического программирования
2. Умение создавать веб-приложения (Django, и т. д.)
3. Уверенное знание JavaScript и CSS

2. Легковесная среда разработки для Erlang
Lightweight Erlang Integrated Development Environment

Разработка законченного программного продукта - среды разработки для языка программирования Erlang с невысокими, по сравнению с JetBrains IDEA, системными требованиями. Функциональные требования: редактирование текста с подсветкой синтаксиса и ошибок компиляции; сборка OTP-проектов; пошаговый отладчик. Опытная эксплуатация программного проекта предполагается в 4-модуле для сопровождения курса "Языки разработки программного обеспечения", если он будет готов к этому сроку.

- Уверенное знание java, основы работы с maven
- Знание или желание изучить API JavaFX
- Крайне желательно посещение курса по выбору "Языки разработки ПО" для понимания назначения и специфики разрабатываемого программного продукта

Яворский Ростислав Эдуардович

Возможные темы курсовых и выпускных квалификационных работ:

Социальные и семантические сети, анализ профессиональных сообществ

[Social and semantic networks, analysis of professional communities]

- Анализ и визуализация коммуникационного графа для блога или форума
Analysis and visualization of the communication graph for a blog or a forum
- Анализ и визуализация данных о научных публикациях по данной теме
Data analysis and visualization of scientific publications on a given topic
- Анализ и визуализация ленты новостей
Analysis and visualization of news feed
- Визуальная аналитика для агрегатора разнородных постоянно обновляемых источников информации
Visual analytics for heterogeneous aggregator of constantly updated sources of information
- *Semantic web* и его применение в разных предметных областях
The Semantic web and its use in different subject areas

Системное программирование: тестирование и верификация программ

[System Programming. Software testing and verification]

- Инструменты статического анализа кода
Tools for static code analysis
- Тестирование распределенных программ и сетевых протоколов
Testing of distributed programs and network protocols
- Верификация ядра операционной системы
Verification of the operating system kernel

Имитационное моделирование

[Simulation]

- Построение модели для выбранной прикладной области (маркетинг, биржевая торговля, биологические системы, ...) и ее анализ
Building a model for the selected application area (marketing, stock trading, biological systems, ...) and its analysis

Темы КР и ВКР

Департамент психологии

2015-2016 учебный год

Список тем **курсовых работ и выпускных квалификационных работ** бакалавров и магистров направления подготовки "Прикладная математика и информатика"

Осадчий Алексей Евгеньевич

aossadtchi@hse.ru

1. Методы неинвазивного функционального картирования мозга (картирование динамики активаций и нейрональных сетей). Методики решения некорректно поставленных задач
Methods of non-invasive functional brain mapping (local and distal synchronization dynamics). Techniques for solving the ill-posed problems
2. Онлайн идентификация систем в задачах создания нейромиографических интерфейсов для управления интеллектуальными протезами
Real-time system identification in neuromyographic interfaces for intelligent control of prosthetic devices
3. Парадигма нейрообратной связи с точки зрения теории обучения с подкреплением. Математическое моделирование
Neurofeedback as a reinforcement learning process. Computer simulations
4. Задачи распознавания образов в интерфейсах мозг-компьютер
Pattern recognition in brain-computer interface technology

Vasily Klucharev/Boris Gutkin/ [Mario Martinez-Saito](#)

Martinez-Saito Mario <selcotsira@hotmail.com>

Моделирование влияния экономической конкуренции на принятие решений. RL и адаптивные стратегии переговоров. Модель на основе исследования *fMRI* принятия решений

Modelling of effect of economic competition on decision making. RL and adaptive bargaining strategies. Model-based fMRI studies of decision making

[Boris Gutkin](#)

bgutkin@hse.ru

Нормативная рациональность защиты гомеостаза: моделирование агентов

A normative rationality of defending homeostasis: agent modelling and simulations

Joseph W. MacInnes

jmacinnes@hse.ru

Вычислительные модели, Eye tracking и исследование зрения

Computational models, Eye tracking and vision research

Attention, Computational Modelling and Eye movement (ACME)

Website <http://social.hse.ru/psy/acme>

<https://sites.google.com/site/wjosephmacinnes/>

Tadamasa Sawada

tsawada@hse.ru

1. Роль в ограничении геодезическими линиями для восприятия поверхности в 3D-сцены
A role of a Geodesics constraint for perceiving a surface in a 3D scene
2. Практические аспекты теории обнаружения сигнала в психофизическом эксперименте
Practical aspects of signal detection theory in a Psychophysical experiment

Zafer Iscan/Vadim Nikulin

zaferiscan@yahoo.com

Алгоритмы классификации

Classification algorithms

Vadim Nikulin

vadim.nikulin@charite.de

Разработка сложных многомерных методик для анализа многоканальных электрофизиологических данных

Development of complex multivariate techniques for the analysis of multichannel electrophysiological data

Презентации проектов департамента психологии Balaev Vladislav

Проект:

Измерение объема области поражения мозга по данным структурной магниторезонансной томографии (МРТ)

Структурная и Функциональная МРТ

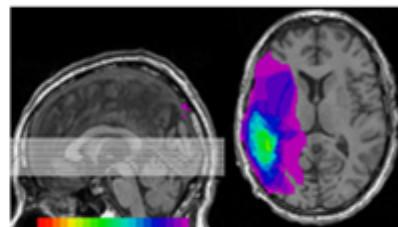
- Зачем: Планирование хирургических операций, радиационной терапии, диагностика заболеваний, планирование приема препаратов, проверка их действия, исследования по неврологии, психологии и т.д.
- Как: Неинвазивно (не нанося вред организму), с помощью сильного магнитного поля и радиочастотного импульса получаем контраст от тканей обладающих различной «восприимчивостью» к магнитному полю
- Функциональная МРТ позволяет отслеживать активации участков мозга по увеличению разницы между количеством оксигенированного и деоксигенированного гемоглобина



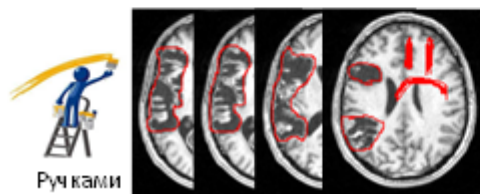
Задача:

Картирование и измерение объема области поражения

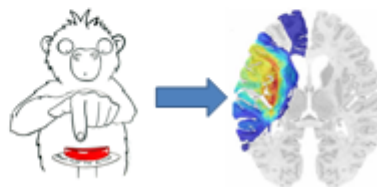
- Зачем: необходимо врачам для планирования операций и/или нейрореабилитационных процедур, приема лекарств и т.п. Для исследователей для определения зависимости характера и степени заболевания от области и объема поражения



Как это делается сейчас:

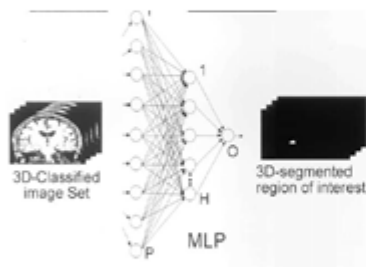


Как хотелось бы :

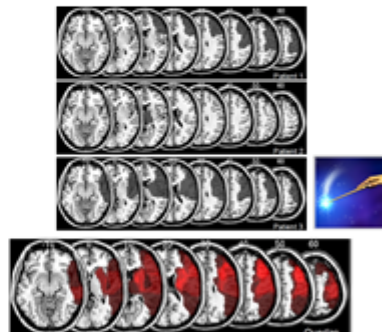


Возможные способы решения

- 1) Автоматически - сложный способ, но наиболее интересный:
- Нейронные сети
- 2) Полуавтоматически, более легкий способ:
- Magic Wand (like in Photoshop)



Или другие алгоритмы схожего типа



Координатор проекта: **Владислав Балаев: vlad_balaev@mail.ru**
Центр нейрoэкономики и когнитивных исследований ВШЭ

A normative rationality of defending homeostasis: agent modelling and simulations

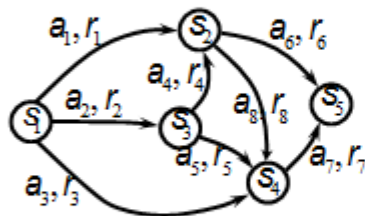
Boris Gutkin, Johann Lussange

HSE Psychology and Group for Neural Theory, ENS, Paris, France

Economic Rationality: Maximize future discounted rewards

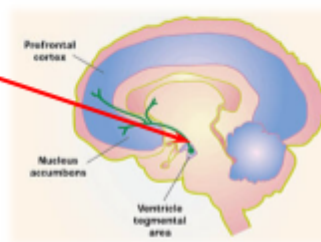
Reinforcement Learning, maximizes reward.

Open issue: What is reward? What is the role of the internal state?



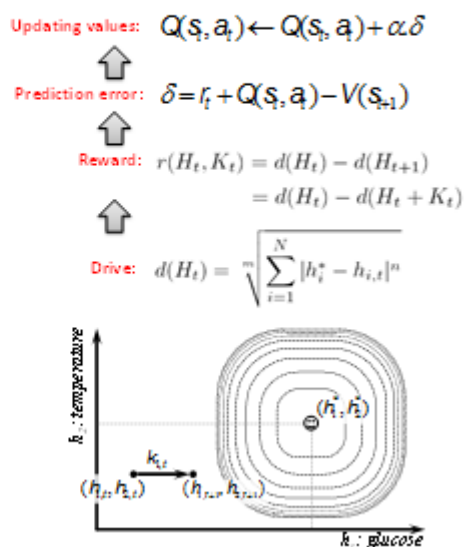
Prediction error: $\delta = r_1 + V(s_2) - Q(s_1, a_1)$

Updating estimates: $Q(s_1, a_1) \leftarrow Q(s_1, a_1) + \alpha \delta$



Homeostatic Reinforcement Learning:

Define Reward as minimizing homeostatic deviation:



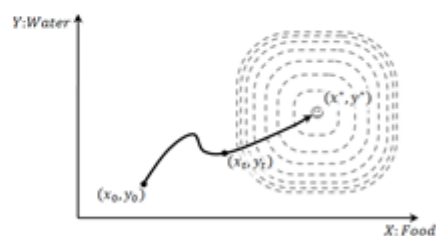
Obtaining reward $\approx$ Defending homeostasis
 Defending Physiological Stability is RATIONAL!

- Any algorithm that maximizes sum of discounted rewards (SDR), also minimizes the sum of discounted drives (SDD), and vice versa.

$\mathcal{P}(H_0)$ is the set of all pathways start at H_0 and end up at H^* .

$$SDD_p(H_0) = \sum_{t=0}^{n-1} \gamma^t d(H_{t+1})$$

$$SDR_p(H_0) = \sum_{t=0}^{n-1} \gamma^t r_t = \sum_{t=0}^{n-1} \gamma^t (d(H_t) - d(H_{t+1}))$$



$$\text{if } \gamma < 1, \quad \underset{p \in \mathcal{P}(H_0)}{\operatorname{argmin}} SDD_p(H_0) = \underset{p \in \mathcal{P}(H_0)}{\operatorname{argmax}} SDR_p(H_0)$$

Project Objective: Explore collective behavior of homeostatically informed and motivated agents

- ▶ Develop agent simulation of at least 2 HRL agents
- ▶ Develop simulated environments with dangerous states
- ▶ Determine under what conditions agents avoid such states
- ▶ Develop simulated environments with high/low resources
- ▶ **Determine agent-agent interactions under various internal states of agents**
- ▶ Explore conditions for cooperative/competitive behavior
- ▶ Methods: agent modeling, machine learning (reinforcement learning), artificial intelligence.

Supervision: Johann Lussange and Boris Gutkin

Computational models, Eye tracking and vision research

Joe MacInnes (Джо МакИннес)

jmacinnes@hse.ru

<https://sites.google.com/site/wjosephmacinnes/>

Office: Room 210

English language lab meetings: every second Friday 4pm

Room 114

Friday 28th 4pm

March 14th 4pm

Overview of research

Topics

Eye Movements and Attention
in complex visual tasks

Visual stability
(construction)

Visual search
(deconstruction)

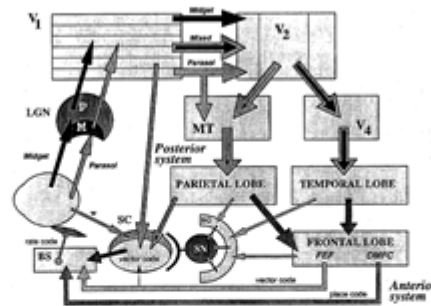
Techniques

Gaze contingent eye tracking

Visual search

Computational models (AI)

Computational models



- Neural model of eye movements From Shiller, 1998
- Create interface that allows easy placement of machine learning 'modules' for each of these components
- This is a project for multiple students
 1. HCVInterface: design a good interface for the overall project
 2. Artificial Intelligence: Develop a computer model of eye movements for one of the modules
 3. Object oriented theory: develop back ground class structure that allows interchangeable AI algorithms

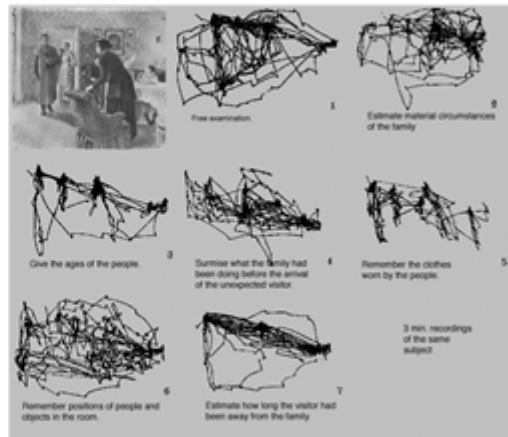
Similarity measure

- **Needleman/Wunsch (NW)(1970):**
 - Sequence similarity measurement used in genetics. Calculates cost to convert one sequence into another (GATGGA -> GGTGTA)
- **Score**
 - +1 = exact match
 - -1 = no similarity



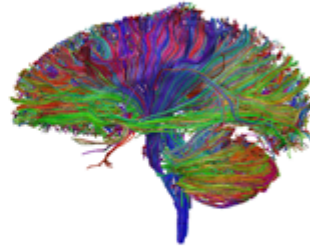
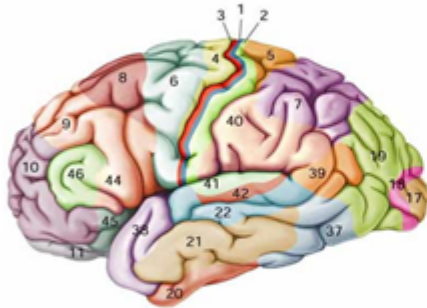
Instructions to saccades

- Yarbus, 1967
- Different instructions yield different patterns of saccades
- Project: develop a way to apply Needleman/Wunsch similarity measure to continuous data like eye movement attributes



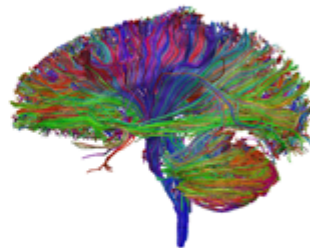
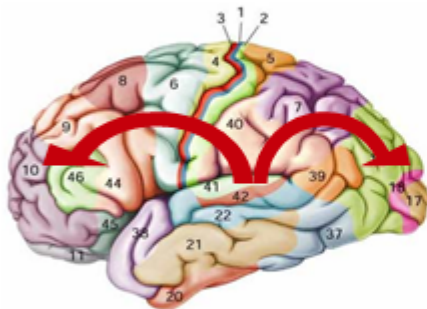
- Thank you!
- Joe MacInnes (Джо МакИннес)
 - jmacinnes@hse.ru
 - <https://sites.google.com/site/wjosephmacinnes/>
- English language Lab Meetings: every second Tuesday 5pm
 - Room 114
 - Tuesday 7th October, 5pm

Brain's anatomy supports cross-regional communication



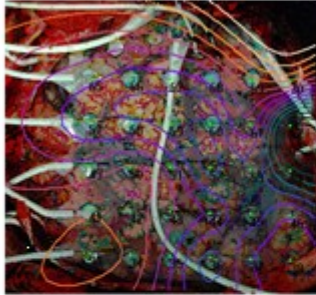
- Brodmann areas originally discovered based on microstructural features were later linked to functions
- Exchange of information between various cortical structures is essential for our cognition

We act when cortical regions talk



- Communication between regions is mediated by neuronal axons and
- Exchange of information between various cortical structures is essential for our cognition

ECoG

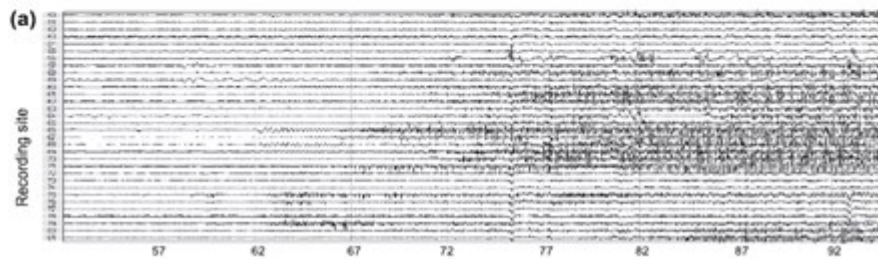


EEG



specialneedsdigest.com

MEG



Dynamic Imaging of Coherent Sources

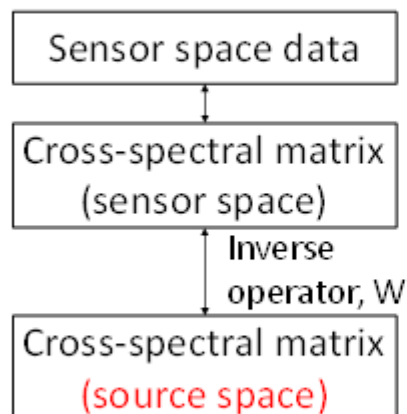
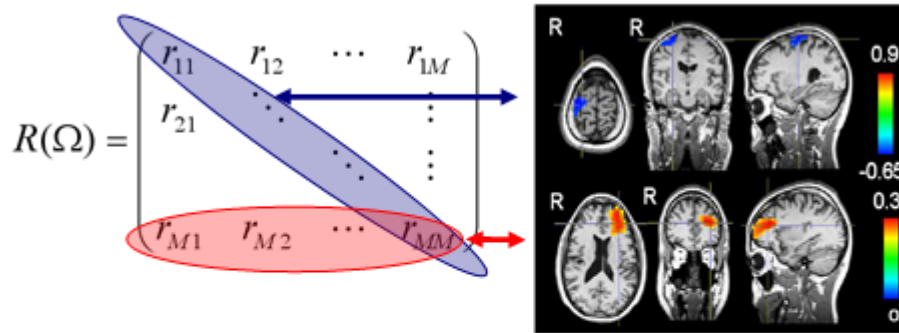


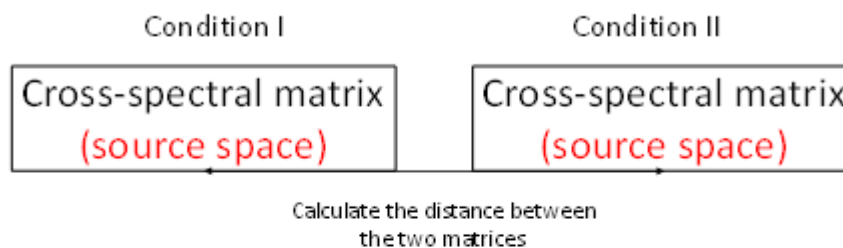
Figure 2: (top) Induced DICS pSPM inverse is shown for selected interval (0.05-0.35 sec, 15-25 Hz). The right central sulcus (arrow) was identified as the first sulcus anterior to the ascending branch of the right cingulate sulcus, which may be identified unambiguously in successive sagittal slices. (bottom) A brain voxel was chosen near the maximum of the event related desynchronization in right primary motor cortex, and coherence was estimated for the remaining brain voxels. The maximum coherence was observed in left dorsolateral prefrontal cortex.



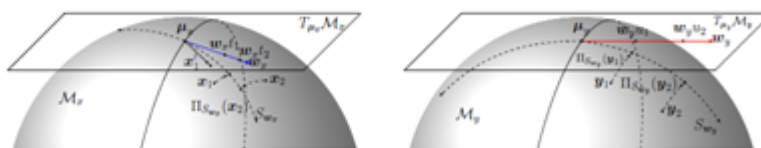
$$R_{source} = W^T R_{sensor} W$$

$$c_{ij} = \frac{r_{ij}}{\sqrt{(r_{ii} r_{jj})}}$$

R.E. Greenblatt, A. Ossadtchi et al, *Front. in Neuroscience Methods*, 2010



Task: Use Log-Euclidean space distance and compare to the simple Euclidean (Frobenius norm based) distance



The use of Riemanian geometry for detection of coupling between cortical regions from non-invasive MEG data

Coordinator: Alexei Ossadtchi, Ph.D.
ossadtchi@gmail.com

Tadamasa Sawada

Faculty of Psychology
Higher School of Economics

tsawada@hse.ru

www.tadamasasawada.com

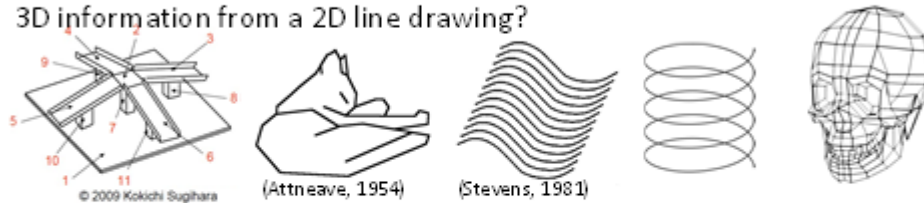
Interests: 3D visual perception, Psychophysics,
Mathematical/Computational modeling

TS1: A role of a Geodesics constraint for perceiving a surface in a 3D scene

TS2: Practical aspects of signal detection theory in a Psychophysical experiment

TS1: A role of a Geodesics constraint for perceiving a surface in a 3D scene

A 3D shape of an object can be well represented by a line drawing with polygonal contours. How the brain (the human visual system) recover 3D information from a 2D line drawing?



Stevens (1981) conjectured that the visual system uses a constraint that the contours are **geodesics** of the surface of the object. However, his idea has been examined only under very limited conditions.

Three definitions of Geodesics

- | | |
|------------------------------|---|
| (1) Locally the shortest | } They lead to the same result
If a surface is twice-differentiable
(Hilbert & Cohn-Vossen, 1952; Koenderink, 1990) |
| (2) Locally the straightest | |
| (3) Locally the most frontal | |

TS1: A role of a Geodesics constraint for perceiving a surface in a 3D scene

Task 1.

Generalizing geodesics for a surface that is not C^2 everywhere and comparing them with the perception.

Methods: Mathematical modeling

Note: A part of this project was already presented (Sawada, ModVis2013).

Task 2.

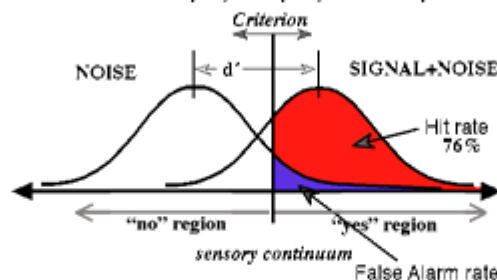
Developing a computational model that takes the Geodesic constraint for recovering a 3D shape of an object from its 2D image and observing how the constraint biases the recovery. The performance of the model will be compared with Psychophysical data.

Methods: Computational modeling (and Psychophysical experiments)

Note: The data will be obtained from existing studies or running the experiments by ourselves

TS2: Practical aspects of signal detection theory in a Psychophysical experiment

Signal detection theory is a way to measure human performance (d') without response bias in a psychophysical experiment.



The theory is formulated under an idealized assumption that number of trials in the experiment is infinite. In this project, we test how the violation of this assumption affects the measured d' and discuss how the effect can be minimized.

Methods: Mathematical and Computational modeling

tsawada@hse.ru

www.tadamasasawada.com