Темы КР и ВКР

Департамент анализа данных и искусственного интеллекта

2015-2016 учебный год

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

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 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

Kertesz-Farkas Attila
Кертес-Фаркаш Аттила
http://www.hse.ru/en/staff/akerteszfarkas
Contact: akerteszfarkas@hse.ru

Projects 1:
Learning to recognize identical molecular biological data using deep learning techniques
[Глубокое обучение для распознания идентичных молекулярных биологических данных]
This project does not require a priori knowledge on biology.
Student in this project will acquire hands-on experience with basic machine learning principles and deep learning techniques (convolution layers, Boltzmann machines, etc); it is a good opportunity to improve his/her English communication, presentation and programming skills.
Level: BSc/MSc

Project 2:
Data acquisition from mobile sensors.
[Сбор данных с мобильных датчиков]
In this project student will implement a program for mobile phone (Android, iPhone, MS), which collects data from the sensors of the mobile phone and then carries out some simple data analysis to recognize user behavior (Running, sleeping, etc). Student in this project will acquire hands-on experience with basic machine learning principles, data analysis methods, and mobile phone developing tools; it is a good opportunity to improve his/her English communication, presentation and programming skills as well.
Level: BSc/MSc

Project 3:
Learning to Rank via Click Through Rate data I
[Ранжирование при помощи обучения на основе выбора пользователя 1]
The aim of this project is to explore the current state-of-the-art methods developed to rank items according to user's preference. This project is being carried out in collaboration with Crossss company (crossss.ru) and student will work on real user generated data along with industrial practitioners.
Student in this project will acquire hands-on experience with basic machine learning principles and learning to rank methods; it is a good opportunity to improve his/her English communication, presentation and programming skills.
Level: BSc/MSc

Project 4:
Learning to Rank via Click Through Rate data II
[Ранжирование при помощи обучения на основе выбора пользователя 2]
The aim of this project is to improve explore the current state-of-the-art methods developed to rank items according to user's preference. This project is being carried out in collaboration with Crossss company (crossss.ru) and student will work with real user generated data and with industrial practitioners.
Student in this project will acquire hands-on experience with basic machine learning principles and learning to rank methods; it is a good opportunity to improve his/her English communication, presentation and programming skills.
Level: MSc

Project 5:
Preference Learning
[Обучение предпочтениям]
This project focuses on learning a preference binary relation. Two objects *x* and *y* are in relation if and only if the *x* is preferred to *y* by a user.
Student in this project will acquire hands-on experience with basic machine learning principles and feature-based preference learning; it is a good opportunity to improve his/her English communication, presentation and programming skills.
Level: BSc/MSc

Project 6:
Multi armed bandits
[Алгоритм многорукого бандита]
Multi-armed bandit problems are receiving a great deal of attention because they adequately formalize the exploration-exploitation trade-off arising in several industrially relevant applications, such as online advertisement and, more generally, recommendation systems. In this project Student will implement and evaluate few current state-of-the-art methods for recommendation systems.
Student in this project will acquire hands-on experience with basic machine learning principles, multi-armed bandits, recommendation systems; it is a good opportunity to improve his/her English communication, presentation and programming skills.
Level: MSc

Project 7:
Fraud detection
[Обнаружение выбросов / мошенничества / ошибок в данных]
The aim of this project is to explore the current state-of-the-art methods developed for fraud detection.
Student in this project will acquire hands-on experience with basic machine learning principles and fraud detection methods; it is a good opportunity to improve his/her English communication, presentation and programming skills.
Level: BSc/MSc

Project 8:
Analyzing human motion data
[Анализ данных передвижения человека]
This project consists of two parts. First acquire human motion data using some wearable 3D human motion tracking wireless devices. The second part is to analyze the data captured and predict signals for missing body parts.
Student in this project will acquire hands-on experience with basic machine learning principles and fraud detection methods; it is a good opportunity to improve his/her English communication, presentation and programming skills.

Project 9:
Clustering user profiles based on their interests [big data]
[Кластеризация профилей пользователей на основе их интересов [большие данные]]
The aim of this project is to group users based on their shopping interests. Features of user are extracted from their browsing history and then student will implement a novel clustering algorithm to group users having similar interests together. This project is being carried out in collaboration with Crossss company (crossss.ru) and student will work with real user generated data and with industrial practitioners.
Student in this project will acquire hands-on experience with basic machine learning principles and unsupervised learning methods; it is a good opportunity to improve his/her English communication, presentation and programming skills.
Level: BSc/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 jacknife. 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 posses 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 do 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
   Automatedreinforcement 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 Semanic Web Techologies

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, Delicioís 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 БМП+)
   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. Онлайновые и распределенные алгоритмы мульмодальной кластеризаци (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 обучения 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 компанией 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 exisiting 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

Макаров Илья Андреевич

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.

<u>Общие темы</u>:

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

Макаров Илья Андреевич

<u>Теоретические темы</u>:

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

а) 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, исследование уже известных алгоритмов и их улучшение с практическим подтверждением проделанной работы. Проблемы сложности алгоритмов и их реализаций
Research of the properties of the algorithm or data structure. Problems of complexity of algorithms and their implementation

(например, инкрементное построение декартового дерева на основе гипотезы о распределении данных)

3. Задачи теории чисел

а) Открытые проблемы многомерных обобщений цепных дробей
Open problems of multidimensional generalizations of continued fractions

(потребуется углубленно изучение работ, связанных с полиэдрами Клейна и гипотезой Оппенгейма с помощью научного руководителя)

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

Макаров Илья Андреевич

<u>Проектные темы КР и ВКР</u>:

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 рефлексы.
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.

Ilya Makarov

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.

Макаров Илья Андреевич

2. Разработка новой методики извлечения знаний на основе данных, полученных от компании «Российские космические системы».

Апробация должна быть представлена в виде разработки проекта стартапа.

Общие требования: умение проектировать сложные системы данных и графически их представлять, работать со сторонним API, навыки методов обработки BigData или визуального распознавания образов или визуализации трехмерной сцены. 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.

Работа осуществляется в рамках поддержки стартапов молодых ученых, аспирантов и студентов компанией «Российские космические системы». Возможно создание проектно-учебной группы по результатам участия в конкурсе с оплатой результатов работы в случае успешной реализации проекта.

Уточняйте конкретные направления исследований у преподавателя.

http://link.springer.com/referencework/10.1007%2F978-3-642-54551-1

Макаров Илья Андреевич

3. 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 неуспеваемос-ти 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 ловарю), 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.

Макаров Илья Андреевич

4. 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 выполения различными студентами заданий по разным типам данных.

5. Свободные темы по инициативным проектам студентов.

Проектные темы для 1 курса:

http://www.ami.hse.ru/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. Методы формирования тематических кластеров в коллекциях текстов: освоение и применение.
   Topic clusters over text collections: methods and applications

2. Методы построения нечетких архетипов и идеальных типов: освоение и сравнение
   Methods of fuzzy clustering involving archetypes and ideal types: coding and comparing

3. Методы иерархического и концептуального кластер-анализа: сравнение и применение.
   Methods for hierarchical and conceptual clustering: comparison and application

4. Построение признаков меньшей гранулярности с помощью кластеров и таксономий.
   Producing features of coarser granularity using clusters and taxonomies

5. Разработка и программирование методов построения аномальных кластеров
   Developing and coding methods for finding anomalous clusters

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 АСНИ «*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.

Объедков Сергей Александрович

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 —> 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 неизбыточность).

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

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 evolving 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. Портирование существующего кода на Си/Си++ в версию, предназначенную для использования в веб-приложении
   Porting an existing C/C++ software code into web-application variant

Разработка программного обеспечения в современном мире ориентирована преимущественно на использование Веб-технологий. Пользователи компьютеров уже воспринимают как данность, что они могут выполнять повседневные задачи с использованием веб-ресурсов, а не прибегая к установке программ на свой компьютер. Тем не менее, имеется огромный пласт наработок в области ПО, как программ, так и библиотек, для которых не существует аналогов, способных работать под управлением веб-браузера. Существующие разработки необходимо либо переписывать на JavaScript, либо использовать полуавтоматизированные способы трансляции в код, который может выполнять браузер.
В предлагаемой курсовой работе предлагается практическое задание по портированию существующего OpenSource проекта (интерпретатора языка программирования), реализованного на языке Си в функционально эквивалентную реализацию на JavaScript. В качестве базового инструмента предлагается использовать Emscripten [https://kripken.github.io/emscripten-site/].

<u>Уверенное знание языков Си и Си++.</u>

2. Реализация многокритериального поиска структурных элементов РНК в базе данных
   Multicriterial search in database of RNA structural elements implementation

Пространственная структура РНК играет ключевую роль в различных внутриклеточных процессах, протекающих с участием РНК. При этом, в отличие от белков, типичные мотивы пространственной структуры РНК не систематизированы и, как следствие, недостаточно изучены. Прежде всего это относится к структурным мотивам, образованным с участием не Уотсон-Криковских взаимодействий (т.н. «неканонические структуры»).

Объектом курсовой работы будет являться аннотированная общедоступная база пространственных структур РНК URS, (Universe of RNA Structures), которая доступна по адресу http://server3.lpm.org.ru/urs/).

Необходимо реализовать часть веб-интерфейса, предназначенную для формирования многокритериальной выборки структурных элементов РНК.

<u>1. Умение работать с базами данных, знание SQL</u>
<u>2. Желание изучать веб-технологии</u>
<u>3. Уверенное владение языком программирования Python и его стандартной библиотекой</u>

Темы ВКР:

1. Исследование структурных мотивов РНК: Псевдоузлы
   RNA structural motifs: Pseudoknots

Пространственная структура РНК играет ключевую роль в различных внутриклеточных процессах, протекающих с участием РНК. При этом, в отличие от белков, типичные мотивы пространственной структуры РНК не систематизированы и, как следствие, недостаточно изучены. Прежде всего это относится к структурным мотивам, образованным с участием не Уотсон-Криковских взаимодействий (т.н. «неканонические структуры»).

Целью работы является выделение в известных пространственных структурах типичных (многократно встречающихся) неканонических структурных мотивов, разработка методик систематизации, классификации и поиска этих мотивов, изучение их роли в функционировании РНК, в частности, при образовании РНК-белковых контактов, изучение эволюции этих мотивов. Основным инструментарием при выполнении работы будет аннотированная общедоступная база пространственных структур РНК URS, (Universe of RNA Structures), которая доступна по адресу http://server3.lpm.org.ru/urs/), а также алгоритмы сравнительного анализа биополимеров. Предлагается провести сравнительный анализ псевдоузлов в экспериментально определенных структурах РНК. Для исследования предлагается использовать топологическую классификацию псевдоузлов, основанную на понятии сигнатуры псевдоузла. Будет проведена разметка сигнатур псевдоузлов для разных значений параметров двойных спиралей, а также проведен анализ зависимости структуры псевдоузла, от имеющихся параметров. Результаты работы будут интегрированы в базу данных URS.

1. Владение базовыми алгоритмами
2. Программирование на Python
3. Понимание устройства нуклеиновых кислот (ДНК, различные РНК)
4. Умение работать с базами данных

Список тем курсовых 2014/2015:

1. Статический анализатор для поиска ошибок в программах на языке Python
2. Регуляторные участки ДНК. Распознавание и оценка статистической значимости.

Проектные темы для 1 курса:

http://www.ami.hse.ru/stud-projects
https://goo.gl/GFbhaq

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",
" Программная инженерия ", "Математика"

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 recogonition 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 ограничения 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

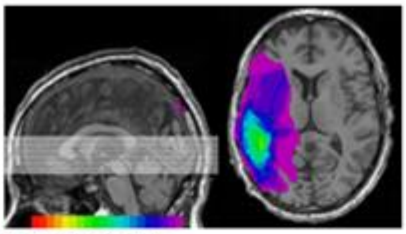
Проект:
Измерение объема области поражения мозга по данным структурной магниторезонансной томографии (МРТ)

Структурная и Функциональная МРТ

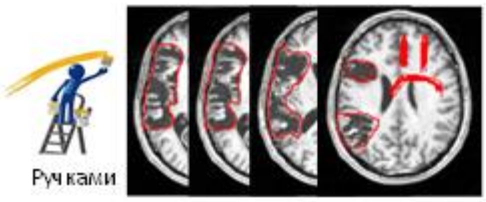
- Зачем: Планирование хирургических операция, радиационной терапии, диагностика заболеваний, планирование приема препаратов, проверка их действия, исследования по неврологии, психологии и т.д.
- Как: Неинвазивно (не нанося вред организму), с помощью сильного магнитного поля и радиочастотного импульса 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

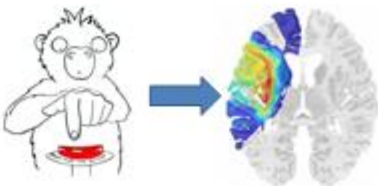
- Зачем: необходимо врачам для планирования операций и/или нейрореабилитационных процедур, приема лекарств и т.п. Для исследователей для определения зависимости характера и степени заболевания от области и объема поражения



Как это делается сейчас:

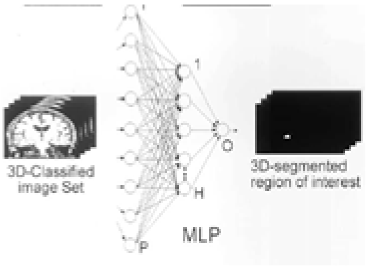


Как хотелось бы :



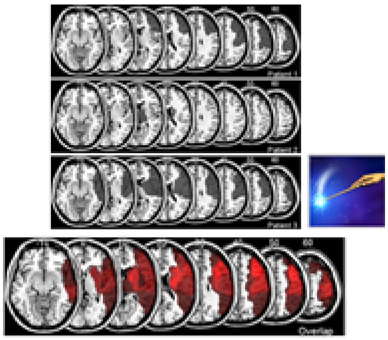
Возможные способы решения

- 1) Автоматически - сложный способ, но наиболее интересный:
- Нейронные сети



Или другие алгоритмы схожего типа

- 2) Полуавтоматически, более легкий способ:
- Magic Wand (like in Photoshop)



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

A normative rationality of defending homeostasis: agent modelling and simulations

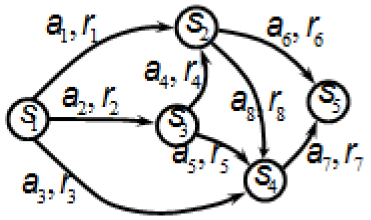
Boris Gutkin, Johann Lussange

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

Economic Rationality: Maximize future discounted rewards

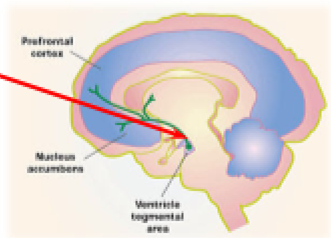
Reinforcement Learning: , maximizes reward.

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:

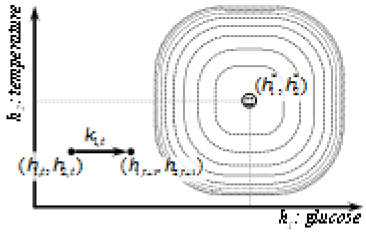
Define Reward as minimizing homoestatic deviation:

Updating values: $Q(s_t, a_t) \leftarrow Q(s_t, a_t) + \alpha\delta$

Prediction error: $\delta = r_t + Q(s_t, a_t) - V(s_{t+1})$

Reward: $r(H_t, K_t) = d(H_t) - d(H_{t+1})$
$= d(H_t) - d(H_t + K_t)$

Drive: $d(H_t) = \sqrt[m]{\sum_{i=1}^{N} |h_i^* - h_{i,t}|^n}$



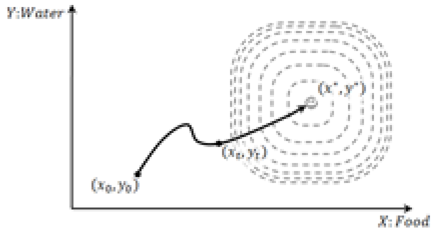
Obtaining reward ≈ Defending homeostasis
Defending Physiological Stability is RATIONAL!

- 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^*

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

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



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

Project Objective: Explore collective behabior 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

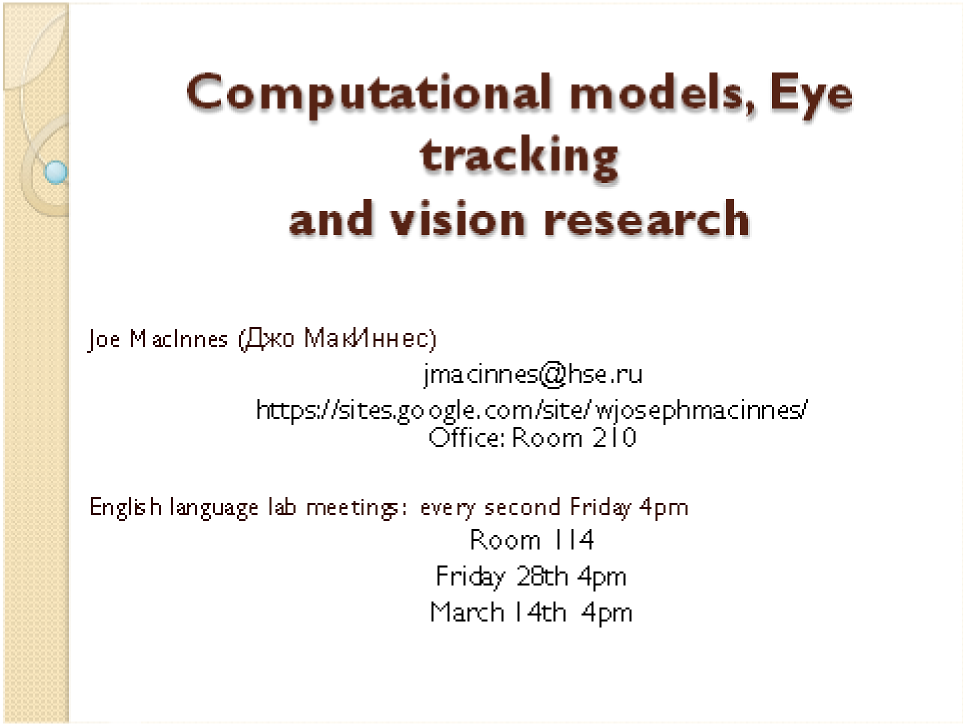
► Develop simulated environments with high/low ressources
►Determine agent-agent interactions under various internal states of agents
► Explore conditions for cooperative/competitive behavior

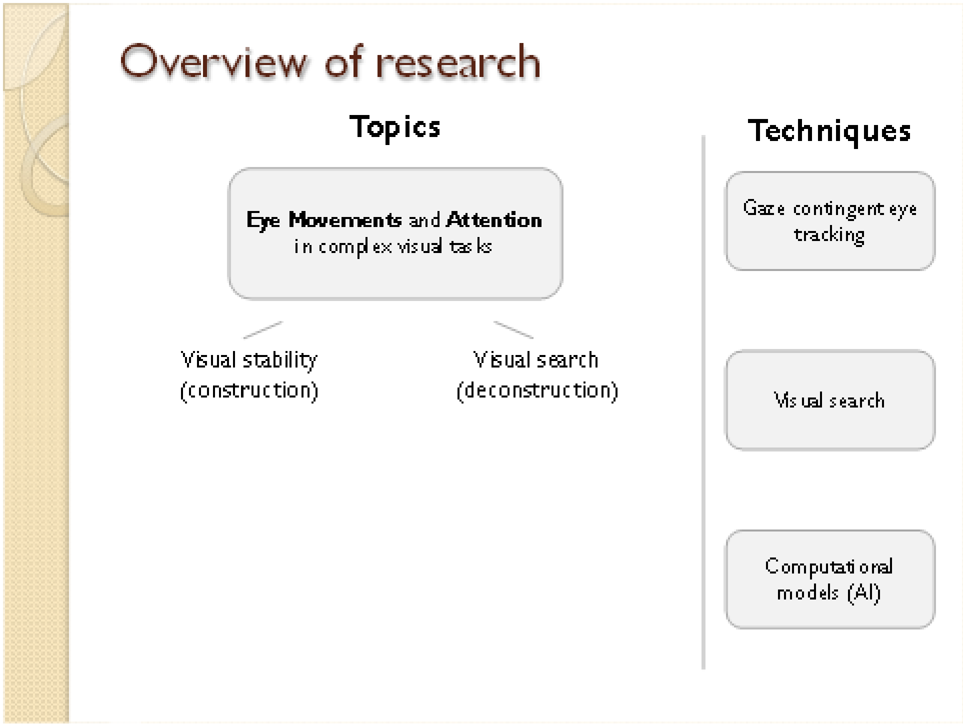
► Methods: agent modeling, machine learning (reinforcement learning), artificial intelegence.

Supervision: Johann Lussange and Boris Gutkin

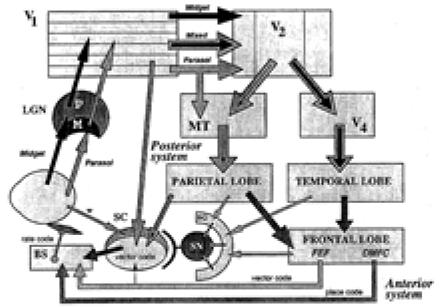
Joe MacInnes

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





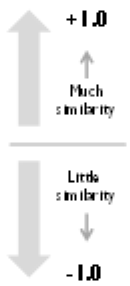
Computational models



- 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. HCI/Interface: 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

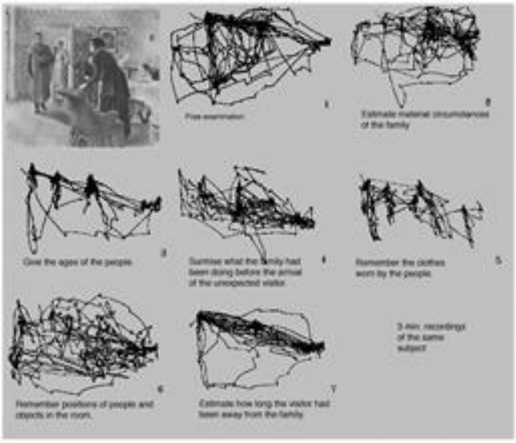
Similarity measure

- **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

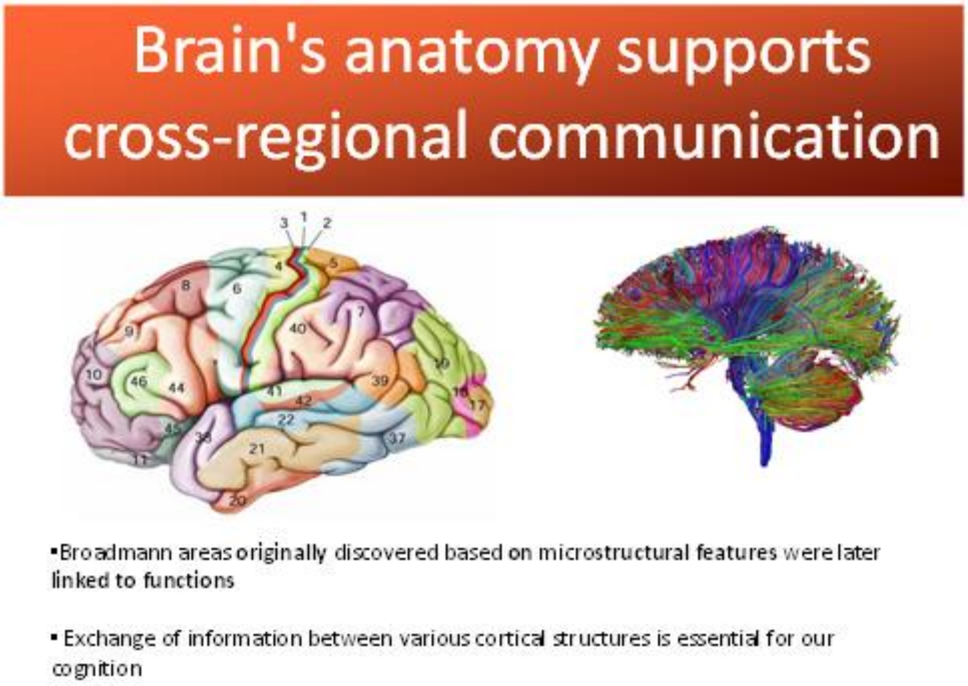
- Yarbus, 1967
- Different instructions yield different patterns of saccades
- Project: develop a way to apply Needleman/Wunsch similarity measure to continuous data like eye movement attributes

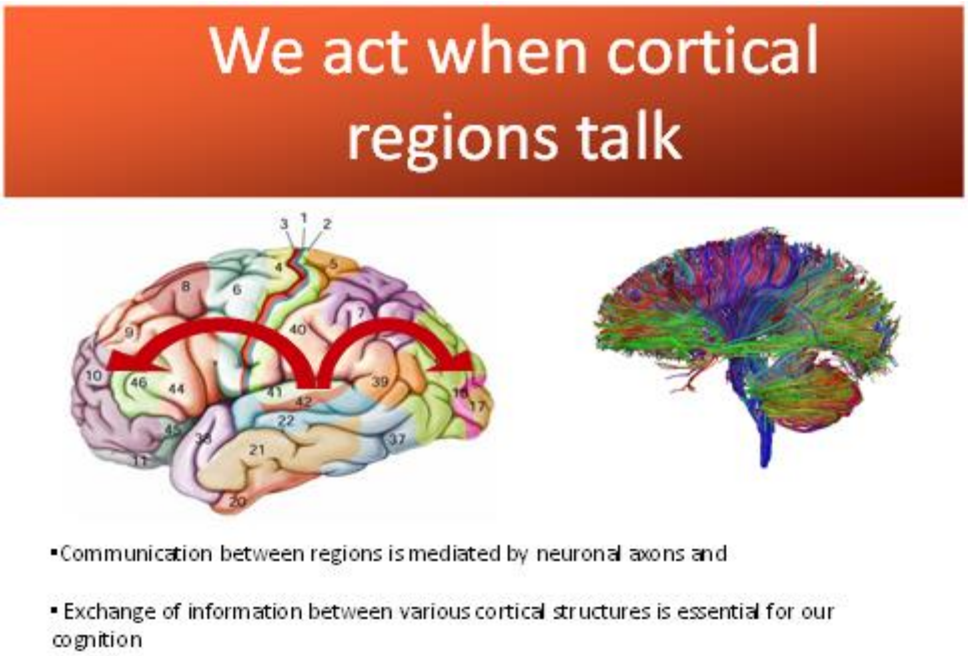


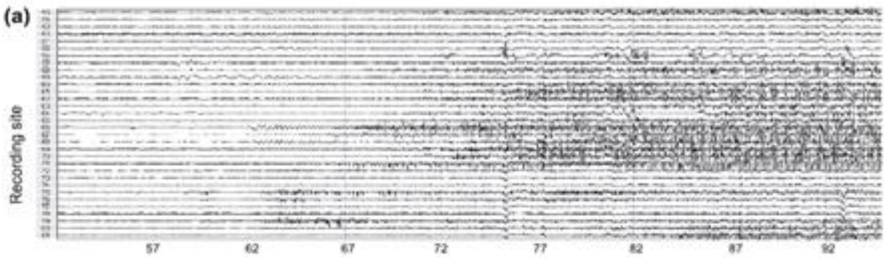
- Thank you!

- Joe MacInnes (Джо МакИннес)
  - jmacinnes@hse.ru
  - https://sites.google.com/site/wjosephmacinnes/
    Office: Room 210

- English language Lab Meetings: every second Tuesday 5pm
  - Room 114
  - Tuesday 7th October, 5pm


Brain's anatomy supports cross-regional communication
▪Broadmann 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



Dynamic Imaging of Coherent Sources

Sensor space data

↕

Cross-spectral matrix (sensor space)

↕ Inverse operator, W

Cross-spectral matrix (source space)

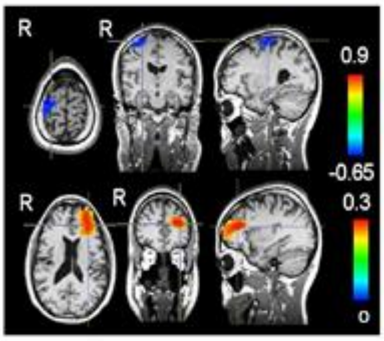


Figure 2: (top) Induced DICS pSPM inverse is shown for selected interval (0.05-0.35 sec, 15-25 Hz). The right central sulcus (arrow) was identified as the first sulcus anterior to the ascending branch of the right cingulate sulcus, which may be identified unambiguously in successive sagittal slices. (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.

$$R(\Omega) = \begin{pmatrix} r_{11} & r_{12} & \cdots & r_{1M} \\ r_{21} & \ddots & & \vdots \\ & & \ddots & \vdots \\ r_{M1} & r_{M2} & \cdots & r_{MM} \end{pmatrix}$$

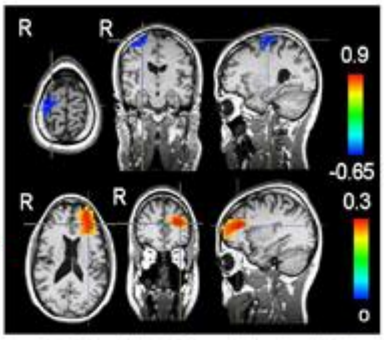


Figure 2: (top) Induced DICS pSPM inverse is shown for selected interval (0.05-0.35 sec. 15-25 Hz). The right central sulcus (arrow) was identified as the first sulcus anterior to the ascending branch of the right cingulate sulcus, which may be identified unambiguously in successive sagittal slices. (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.

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

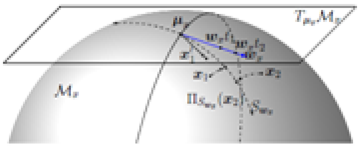
$$c_{ij} = \frac{r_{ij}}{\sqrt{(r_{ii} r_{jj})}}$$

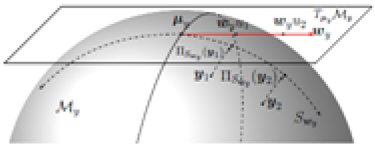
R.E. Greenblatt, A. Ossadtchi et al, Front. in Neuroscience Meths., 2010

Condition I	Condition II
Cross-spectral matrix (source space)	Cross-spectral matrix (source space)

Calculate the distance between the two matrices

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

Sawada Tadamasa

Tadamasa Sawada
Faculty of Psychology
Higher School of Economics

tsawada@hse.ru
www.tadamasasawada.com

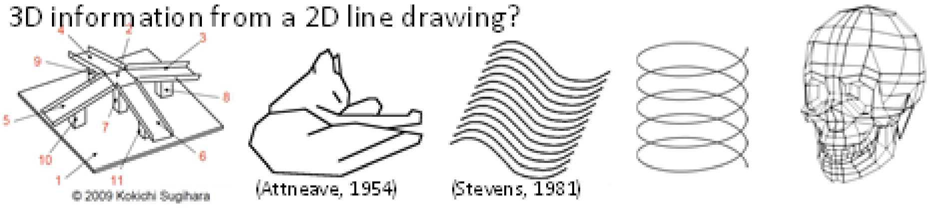
Interests: 3D visual perception, Psychophysics, Mathematical/Computational modeling

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
(2) Locally the straightest
(3) Locally the most frontal

They lead to the same result
If a surface is twice-differentiable
(Hilbert & Cohn-Vossen, 1952; Koenderink, 1990)

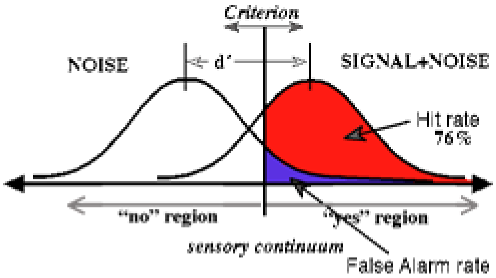
TS1: A role of a Geodesics constraint for perceiving a surface in a 3D scene

Task 1.
Generalizing geodesics for a surface that is not *C2* 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

Темы КР и ВКР

Бывшие преподаватели, аффилированные с Департаментом АДИИ

2015-2016 учебный год

Написание КР или ВКР возможно в рамках назначения научного консультанта от департамента

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
hse.ru/staff/iamakarov

Ilya Makarov
hse.ru/en/staff/iamakarov

К каждому преподавателю приложены 1-2 ссылки:
Ссылка на персональную страницу (гиперссылка)
Ссылки на доп. материалы

Некоторые КР и ВКР ранжированы по разделам, которые должны быть указаны при презентации студентам.

The following projects are considered:

Project 1: Semi-automated keyword extraction

Traditional fully automated keyword extraction methods including tf-idf have low performance when they are used to extract keywords and key phrases from short texts. In this project students will investigate the performance of semi-automated extraction techniques. Starting from a list of keywords provided by domain experts, co-occurrence networks can be used to identify other words and phrases which are often mentioned in the same documents. A first challenge will be to filter out common phrases and only retain domain specific ones. A second challenge consists of identifying proper measures for annotating and pruning these networks. A simple example is the relative frequency.

Project 2: Including human expertise in self-organising maps

Self-organising maps are a powerful instrument for analyzing high-dimensional unsupervised data. The maps can be used to identify interesting patterns in such data. However, human expertise is often not included resulting in suboptimal visualizations. The topic of this project is to incorporate such human expertise by for example using weights assigned by experts to features in the training process. A second task would be to develop intelligent "coloring" methods which effectively guide the human expert in his discovery task. For example, the user may have a particular interest in certain sub populations with particular features. Intelligent highlighting and offering facilities for convenient in depth exploration of areas of interest should also be provided.

Project 3: Text mining scientific papers on care pathways and care process analysis

This project is an applied research work where the focus is not on developing new methods for text mining but optimizing the facilities which are provided in the CORDIET framework. In particular, the student will focus on building a domain specific thesaurus. After collecting relevant papers, these papers will be indexed with CORDIET and concept lattices can be generated that give an overview of the literature. In a last phase, these lattices and references to publications can then be made public through a web portal where medical practitioners can use them to easily retrieve relevant publications

Project 4: Performance comparison of Gaussian Mixture Models with other state of the art techniques for classifying textual documents.

Gaussian Mixture Models are a recent technique for classifying high dimensional non-homogeneous data. Till date there are no convincing applications in the text mining and text classification field. This project will consist of comparing the performance of this method with approaches such as support vector machines and neural networks. Experimental validation can be done on the Reuters benchmarking dataset.

Борисенко Владимир Витальевич

Написание КР или ВКР возможно в рамках назначения научного консультанта от департамента

Список тем можно прочитать также в Интернете по адресу:
http://mech.math.msu.su/~vvb/HSE/topics.htm

Любая тема может быть, как курсовой работой, так и магистерской, в зависимости от уровня исполнения и получения собственных оригинальных результатов.

1. Надежные (Robust) методы вычисления статистических показателей, основанные на понятии медианы.
2. Способы публикации документов, содержащих математические выражения, в сети Internet.
3. Алгоритмы поиска поддеревьев и их применение к поиску математических выражений, представленных в формате MathML.
4. Реализация графического векторного редактора, использующего язык PostScript.
5. Обзор криптографических алгоритмов, основанных на применении теории чисел.
6. Алгоритмы поверхностной, объемной и смешанной визуализации в трехмерной графике.
7. Алгоритмы интерполяции контуров на плоскости и поверхностей в трехмерном пространстве, заданных триангуляциями.
8. Реализация методов математической морфологии для трехмерных объектов.
9. Алгоритмические проблемы эквивалентности конечных автоматов и регулярных выражений.
10. Минимизация недетерминированных конечных автоматов.

Гребенюк Елена Алексеевна

Написание КР или ВКР возможно в рамках назначения научного консультанта от департамента

Темы КР и ВКР 2014

1.Оценивание эффективности применения нового лекарственного средства при лечении конкретного заболевания.
2. Сравнение выживаемости в двух группах пациентов.
(3-4 курс)

Курсовые и магистерские курсовые

1. Массовая оценка кредитных рейтингов заемщиков.
2. Анализ влияния основных макроэкономических факторов на цены на недвижимость.
3. Моделирование стоимости жилья на рынке новостроек. КРМаг.
4. Применение пространственной регрессии для моделирования рынка недвижимости.
5. Модели стохастической волатильности. Их свойства и сравнение.. КРМаг
6. Анализ финансового состояния группы предприятий, формирование рейтинговых оценок. КР
7. Методы массовой оценки объектов недвижимости. КР
8. Алгоритмы построения виртуальных анализаторов и их применение для контроля производственных процессов. КРМаг
9. Применение методов фильтрации в задачах мониторинга

ВКР

1. Применение методов фильтрации для краткосрочного прогнозирования финансовых и макроэкономических показателей
2. Модели финансовых пузырей и методы их идентификации
3. Построение массовой оценки объектов недвижимости с учетом влияния местоположения объектов на цену
4. Прогнозирование цунами с использованием методов последовательного анализа данных
5. Влияние экологической обстановки и климата на здоровье населения.
6. Анализ ЭЭГ на основе выделения стационарных сегментов

Корухова Юлия Станиславовна

Написание КР или ВКР возможно в рамках назначения научного консультанта от департамента

Автоматическое доказательство логических теорем.

Курсовая работа посвящена исследованию методов автоматического доказательства теорем. В качестве базового метода доказательства рассматривается метод резолюций. Базовый метод включает в себя набор правил вывода, однако для его применения на практике необходимо дополнительно предложить стратегию применения правил. В рамках работы предлагается разработать систему, позволяющую ввести условие задачи - теоремы для доказательства, выбирать стратегию применения правил (запросить ее название у пользователя либо сделать выбор автоматически, используя эвристики) и наблюдать за построением доказательства. В режиме, когда стратегия выбирается из заранее заданного списка, система становится помощником пользователя в изучении метода резолюций. Для автоматического подбора стратегии, нужно предложить эвристики, позволяющие оценить, какая последовательность правил лучше подойдет для построения доказательства.

Разработка туристической рекомендательной системы.

Курсовая работа посвящена исследованию методов построения рекомендаций и применению таких методов для составления планов экскурсий или туристических поездок. Исходными данными системы являются отзывы путешественников, хранящиеся в сети Интернет в открытом доступе, и информация, полученная от пользователя системы: возраст, длительность поездки, сумма денег, предпочитаемый вариант отдыха – спорт, пляжный отдых и другие варианты. В качестве результата работы системы ожидается одна или несколько программ отдыха, упорядоченных по уменьшению релевантности запросу.

Лукашевич Наталья Валентиновна

Написание КР или ВКР возможно в рамках назначения научного консультанта от департамента

Темы КР, ВКР и магистерских

Работы для студентов всех уровней, будет видоизменяться сложность работы.
Важно умение программировать.

1. Автоматическое извлечение из текстовой коллекции наиболее частотного значения слова для тезауруса русского языка РуТез (http://labinform.ru/ruthes/index.htm)
   (Automatic extraction of the most frequent sense of an ambiguous word)

Многозначность слов является фактором, который значительно усложняет автоматическую обработку текстов. При этом для многих многозначных слов известно, что в большинстве случаев они употребляются в одном из значений. В работе предполагается проверить несколько алгоритмов для получения наиболее частотных значений многозначных слов.

2. Автоматический анализ тональности текстов по отношению к заданному объекту (персоне, организации).
   (Entity-oriented sentiment analysis)

Тональность текста по отношению к заданному объекту может основываться на двух различных факторах:
– отрицательное, положительное или нейтральное отношение автора текста к заданному объекту или
- отрицательный, положительный или нейтральный факт по отношению к заданному объекту.
Анализ тональности текста представляет собой автоматическую классификацию текстов на три категории: отрицательный, положительный или нейтральный в зависимости от тональности текста по отношению к заданному объекту.

3. Современные методы извлечения из текстов именованных сущностей.

(Modern approaches to named entity recognition)

Для извлечения информации из текстов нужно, прежде всего, точно и полно извлекать названия конкретных сущностей (имен людей, организаций, географических мест).
Работа может иметь два варианта:
3.1. Накопление словарей и правил для извлечения имен,
3.2. Построение специальных текстовых коллекций для последующего применения методов машинного обучения

4. Автоматическая тематическая структуризация текстовой коллекции

(Topical structuraztion of text collections)

При обработке больших коллекций текстов часто встает задача проанализировать тематический состав таких коллекций. Одним из популярных подходов к такому тематическому анализу является применение специальных вероятностных тематических моделей. В работе предполагается исследовать сочетание известных статистических методов с различной лингвистической информацией.