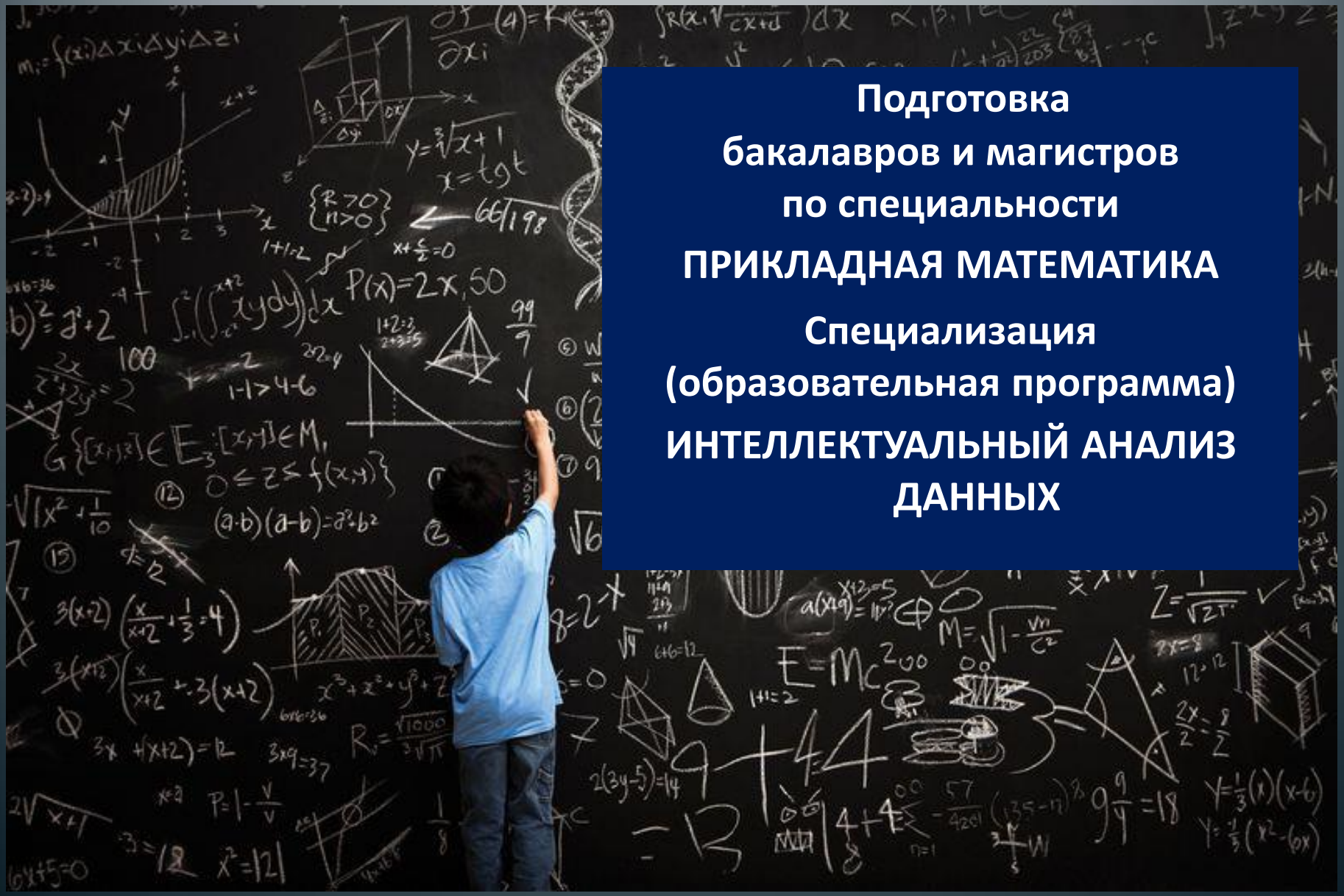
НТУ «ХПИ»

Кафедра «Компьютерная математика и анализ данных»

АНАЛИЗ И ИНЖЕНЕРИЯ ДАННЫХ - СПЕЦИАЛЬНОСТЬ БУДУЩЕГО

Проф. Л.М. Любчик

Кафедра
"КОМПЬЮТЕРНАЯ МАТЕМАТИКА И АНАЛИЗ ДАННЫХ"



Подготовка

бакалавров и магистров

по специальности

ПРИКЛАДНАЯ МАТЕМАТИКА

Специализация

(образовательная программа)

ИНТЕЛЛЕКТУАЛЬНЫЙ АНАЛИЗ ДАННЫХ

Наука о данных (Data Science)

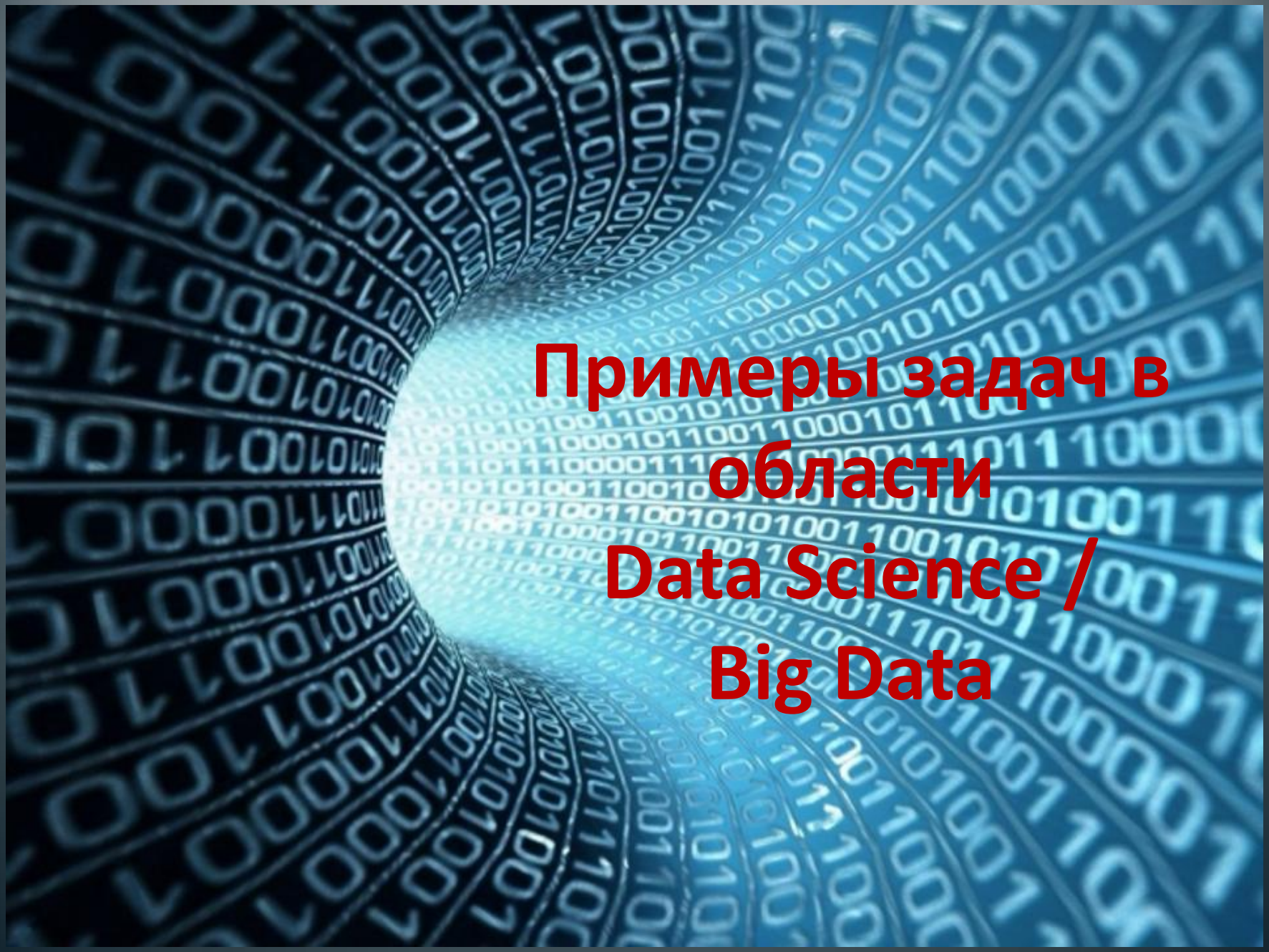
- **Наука о данных (*Data Science*) – раздел информатики, связанный с проблемами анализа, обработки и представления данных с целью извлечения знаний и прогнозирования.**
- **Объединяет статистические методы, методы и приложения машинного обучения и искусственного интеллекта, а также методы обработки больших массивов данных на основе информационных технологий распределенных и облачных вычислений.**

Что должен знать и уметь специалист по данным?

- Знать методы машинного обучения (Machine Learning), вычислительного интеллекта (Computational Intelligence) и мягких вычислений (Soft Computing)
- Уметь осуществлять информационный поиск и извлечение знаний (Data Mining)
- Обладать навыками разработки и эксплуатации программных средств обработки сверхбольших массивов данных (Big Data), технологиями облачных вычислений (Cloud Computing).

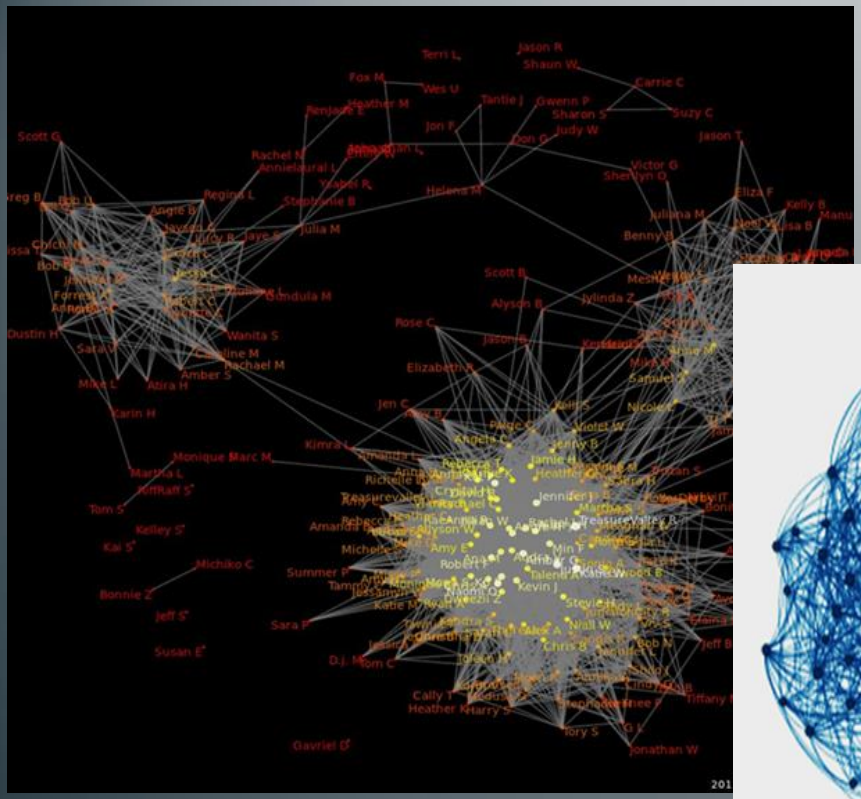
«Большие данные» (Big Data)

- **Большие данные (*Big Data*) совокупность подходов, инструментов и методов обработки неструктурированных данных огромных объёмов и значительного многообразия для получения воспринимаемых человеком результатов.**
- ***«Три V»*: объём (*volume*), скорость (*velocity*), многообразие (*variety*).**
- **Базовые технологии: NoSQL, MapReduce, Hadoop**

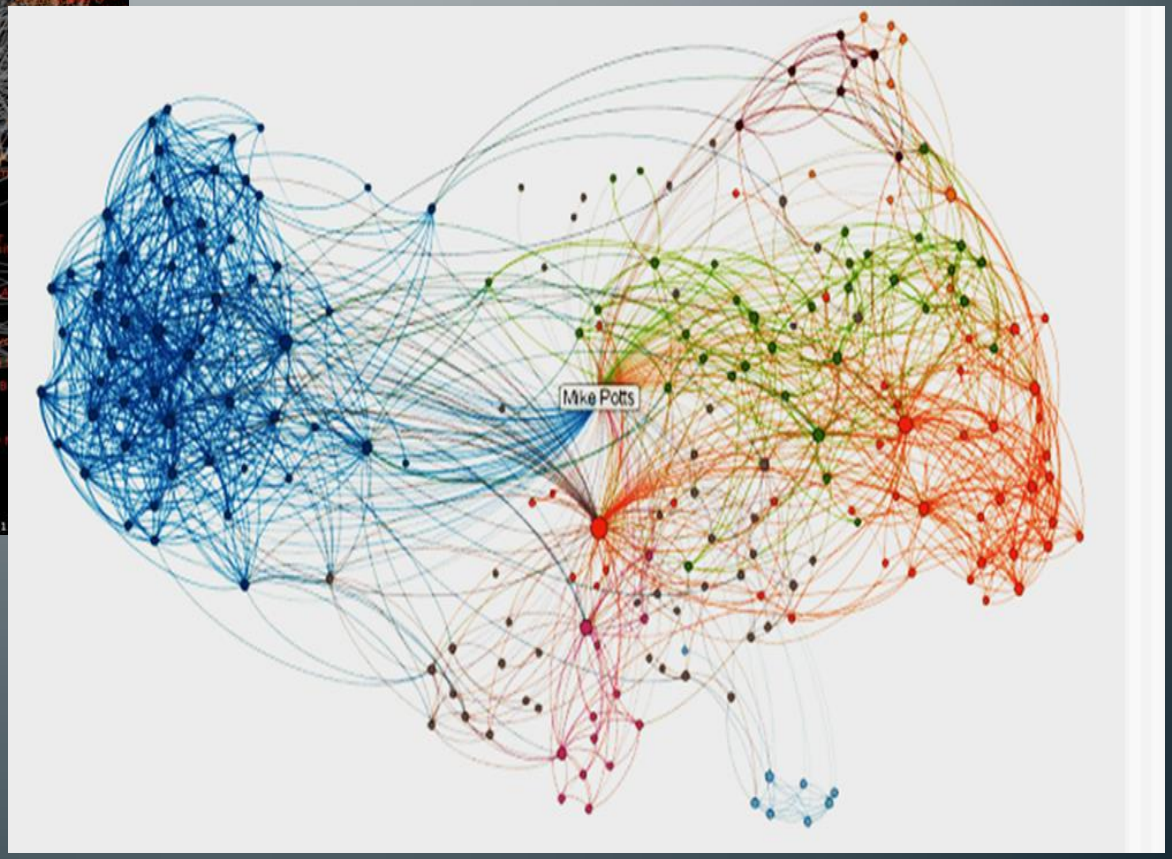


Примеры задач в области Data Science / Big Data

Модели социальных сетей

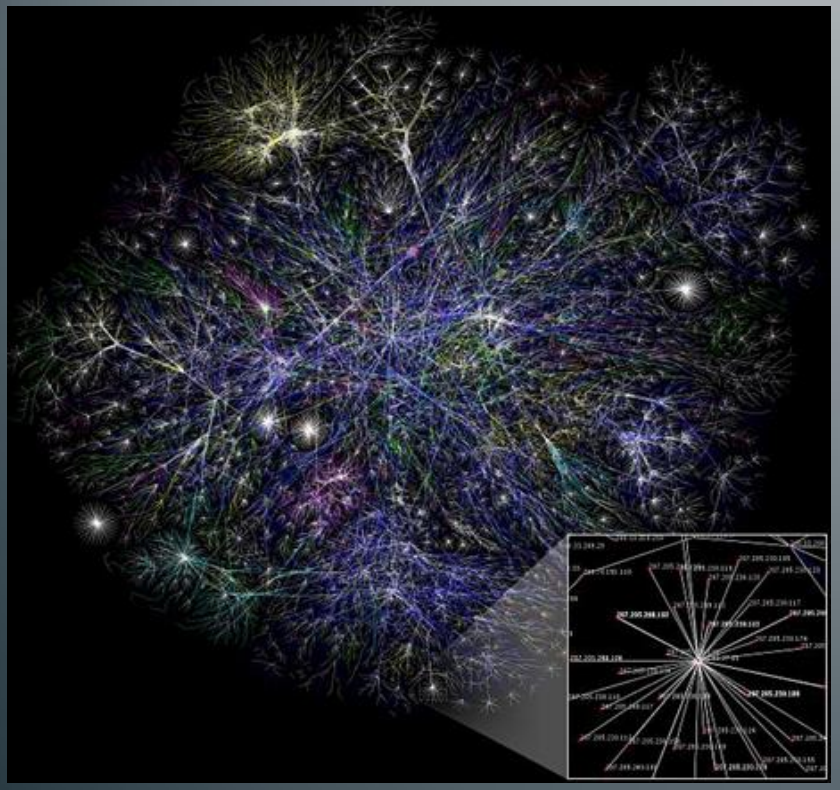


Сеть пользователей Facebook

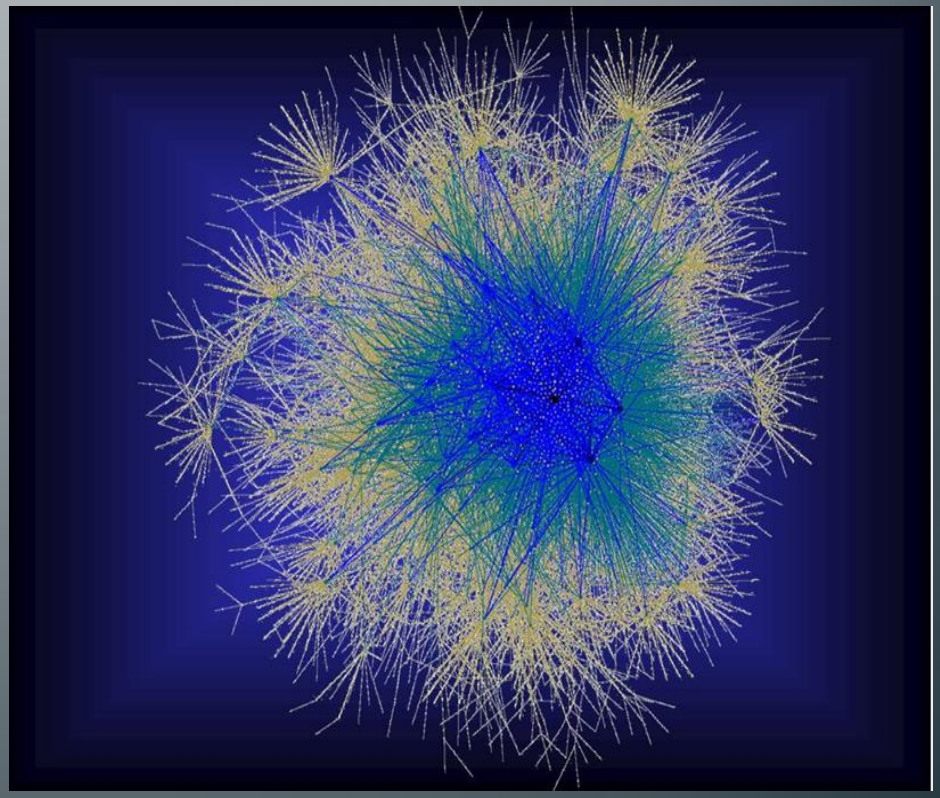


Структура профессиональной социальной сети LikendIn

Визуализация и анализ графовых моделей WEB - пространства

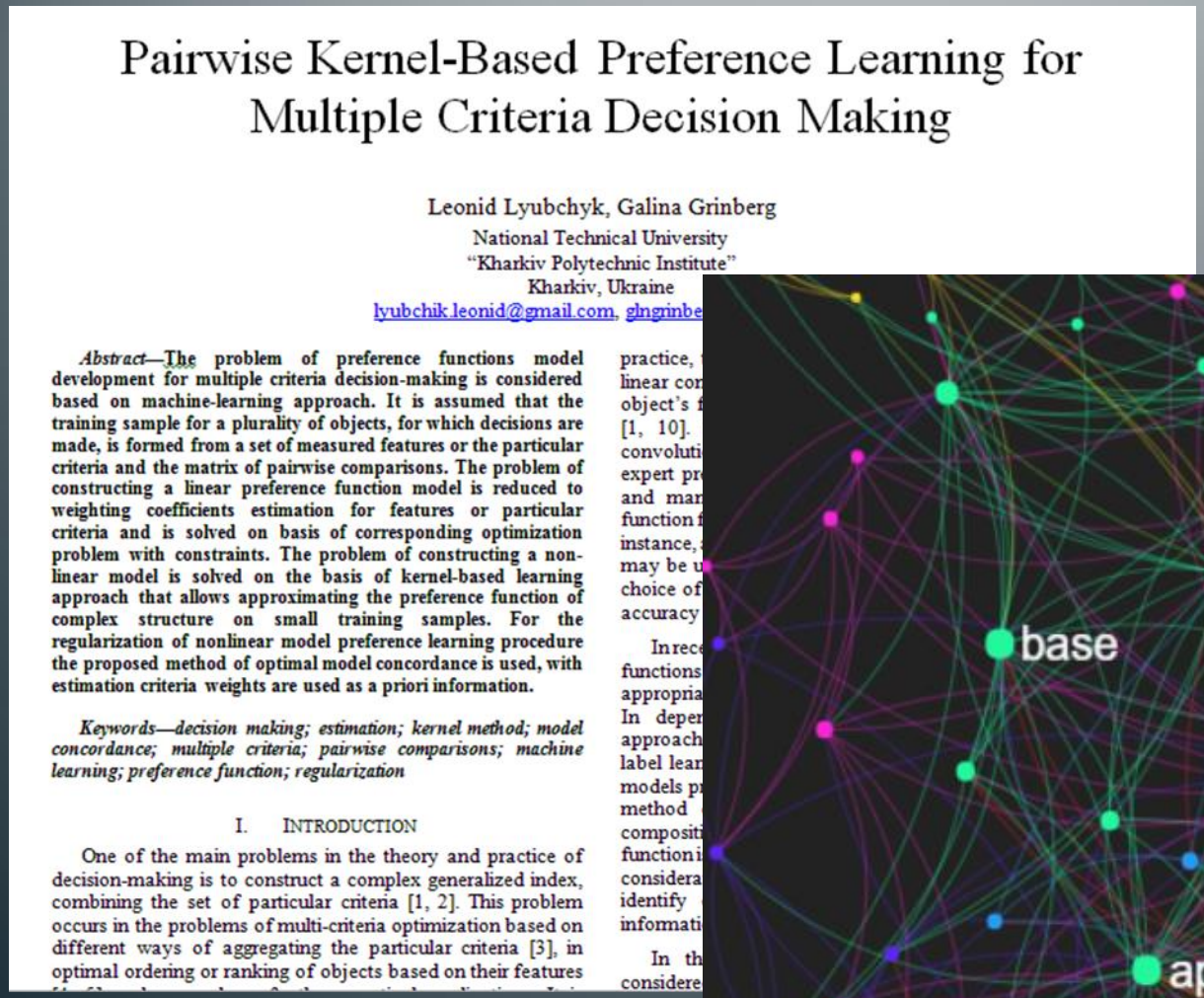


Структура связей между сайтами



Веб -граф

Построение сетевой модели структуры текста



Pairwise Kernel-Based Preference Learning for Multiple Criteria Decision Making

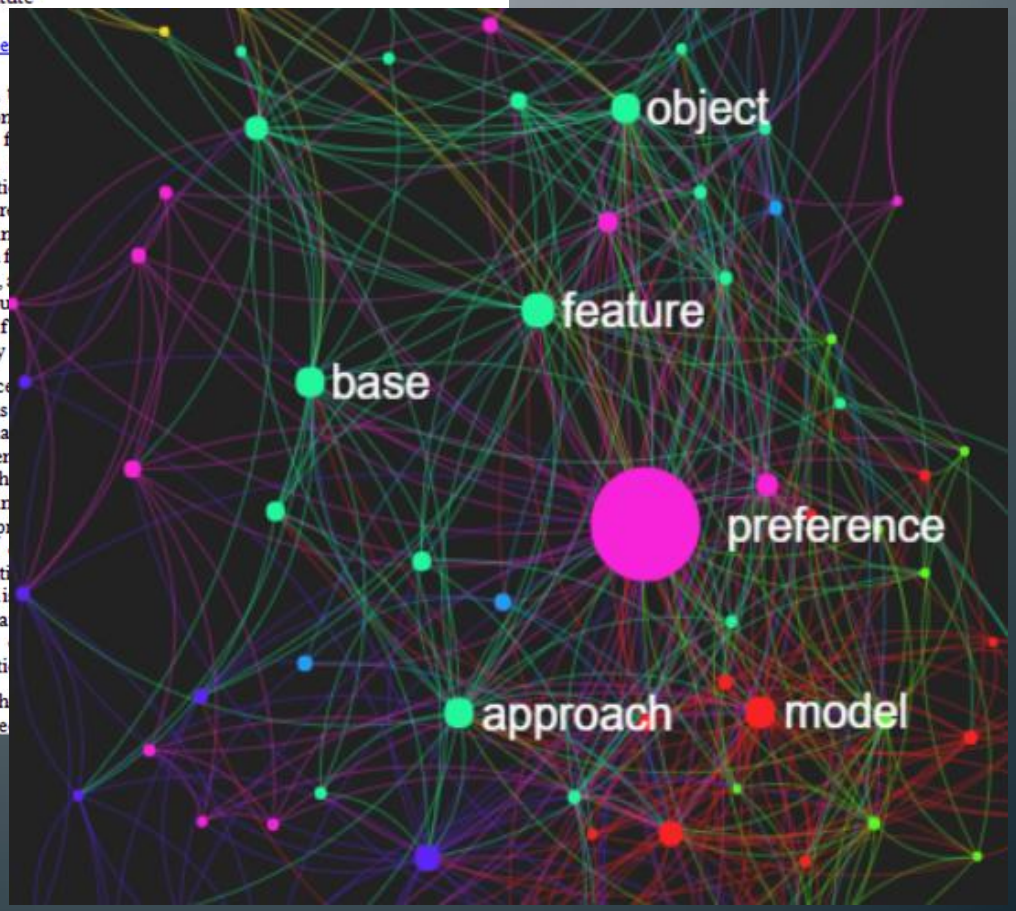
Leonid Lyubchyk, Galina Grinberg
National Technical University
"Kharkiv Polytechnic Institute"
Kharkiv, Ukraine
lyubchik.leonid@gmail.com, glngrinbe

Abstract—The problem of preference functions model development for multiple criteria decision-making is considered based on machine-learning approach. It is assumed that the training sample for a plurality of objects, for which decisions are made, is formed from a set of measured features or the particular criteria and the matrix of pairwise comparisons. The problem of constructing a linear preference function model is reduced to weighting coefficients estimation for features or particular criteria and is solved on basis of corresponding optimization problem with constraints. The problem of constructing a non-linear model is solved on the basis of kernel-based learning approach that allows approximating the preference function of complex structure on small training samples. For the regularization of nonlinear model preference learning procedure the proposed method of optimal model concordance is used, with estimation criteria weights are used as a priori information.

Keywords—decision making; estimation; kernel method; model concordance; multiple criteria; pairwise comparisons; machine learning; preference function; regularization

I. INTRODUCTION

One of the main problems in the theory and practice of decision-making is to construct a complex generalized index, combining the set of particular criteria [1, 2]. This problem occurs in the problems of multi-criteria optimization based on different ways of aggregating the particular criteria [3], in optimal ordering or ranking of objects based on their features



Кафедра НТУ «ХПИ» "КОМПЬЮТЕРНАЯ МАТЕМАТИКА И АНАЛИЗ ДАННЫХ"

Кафедра обеспечивает подготовку бакалавров и магистров по специальности 113 - «Прикладная математика» со специализацией в области математических методов анализа данных.

Выпускники кафедры, владеющие современными методами прикладной математики и компьютерных наук, готовятся для научной и производственной деятельности в области разработки и применения математических моделей, информационных технологий и программного обеспечения для анализа данных, извлечения знаний, управления и принятия решений.

Специальность "Прикладная математика“ с фокусом в анализ данных

- фундаментальное математическое образование – прочная основа будущей научной и производственной карьеры
- профессиональная компьютерная подготовка – гарантия трудоустройства в Украине и за рубежом
- специальная прикладная математическая и компьютерная подготовка в области методов и технологий моделирования, принятия решений и анализа данных – база для работы в сфере исследований и разработок (R&D)

Учебный план специальности ПМ на кафедре КМАД

1	2	3	4	5	6	7	8
Фізичне виховання 2	Фізичне виховання 2	Фізичне виховання 2	Фізичне виховання 2	Фізичне виховання 2	Фізичне виховання 2		
Іноземна мова 2	Іноземна мова 2	Іноземна мова 2	Іноземна мова 2				
Українська мова 3	Історія української культури 2	Філософія 3	Історія науки і техніки 2	Правознавство 2	Политология 2	Социология 2	Психология 2
Історія України 3				Основи економічної теорії 2	Основи Екології 2	Професійна безпека і здоров'я людини 3	Економіка виробництва 3
Математический анализ 6	Математический анализ 6	Математический анализ 6	Дифференциальные и разностные уравнения 4	Уравнения в частных производных 3	Теория управления 4	Теория принятия решений 3	Нечеткие и мягкие вычисления 3
Линейная алгебра 5	Линейная алгебра 5	Теория вероятностей 4	Математическая статистика 4	Анализ данных 4	Временные Ряды 4	Машинное обучение 4	Представление и извлечение знаний 3
			Функциональный анализ 4	Численные методы 4	Методы оптимизации 4	Параллельные вычисления 4	Интеллектуальный поиск 2
	Математическая логика 4	Дискретная математика 4	Компьютерная геометрия и графика 4	Случайные процессы 4	Обработка сигналов и изображений 4	Анализ и обработка текстов 3	Семантический WEB 2
Программирование C++ 5	Программирование C++, Pyton 5	Объектно-ориентированное программирование 5	Структуры данных 5	Теория алгоритмов 5	Технологии программирования 4	Облачные технологии программирования 4	
Программное обеспечение 4	Архитектура компьютеров 4	Компьютерные сети 4	Математическое и компьютерное моделирование 3	Базы данных 4	Проектирование информационных систем 4	Проектирование программного обеспечения 4	Преддипломная практика 6
						Нейросетевые технологии 3	Дипломное проектирование 6 Аттестация 3
	Курсовая работа П+МЛ	Курсовая работа ООП+ДМ	Курсовая работа СД + КГГ	Курсовая работа ТА + ЧМ	Курсовой проект ПИС	Курсовой проект ППО	Дипломная работа

Образовательная программа (специализация) «Интеллектуальный анализ данных» на кафедре «Компьютерной математики и анализа данных»

- Образовательная программа направлена на подготовку специалистов, владеющих современными математическими методами и информационными технологиями интеллектуального поиска, анализа, обработки и визуализации данных с целью извлечения знаний и прогнозирования.
- Цель подготовки по образовательной программе заключается в овладении методами машинного обучения, мягких вычислений и вычислительного интеллекта, умениями и навыками разработки и эксплуатации программных средств обработки сверхбольших массивов данных на основе технологий распределенных и облачных вычислений.

Образовательная программа (специализация) «Интеллектуальный анализ данных» на кафедре «Компьютерной математики и анализа данных»

Профессиональная деятельность выпускников связана

- с разработкой алгоритмов и программ обработки неопределенных и неструктурированных данных применительно к задачам анализа текстов, сигналов и изображений, поиска и извлечения знаний;

- с участием в разработке проектов, связанных с классификацией и ранжированием Web-объектов (клиентов, пользователей и сайтов), моделированием поисковых систем, прогнозированием трафика, анализом и моделированием взаимодействия клиентов и сервисов, выявлением и идентификацией аномалий и угроз.

Структура подготовки по специализации "Интеллектуальный анализ данных” на квалификационном уровне «Бакалавр»

- методы машинного обучения,
- методы извлечения и представления знаний,
- обработка сигналов и изображений,
- визуализация данных,
- методы защиты информации,
- математическая лингвистика,
- компьютерный анализ и обработка текстов,
- информационный поиск, семантический WEB.

Структура подготовки по специализации "Интеллектуальный анализ данных« на квалификационном уровне «Магистр»

- распределенные и корпоративные информационные системы;
- параллельные вычисления и облачные технологии;
- нечеткие модели и методы, мягкие вычисления, нейросетевые технологии; основы вычислительного интеллекта;
- экспертные системы и базы знаний;
- интеллектуальный поиск и анализ данных в Интернете.

Научные, прикладные и коммерческие разработки кафедры КМАД в области Data Science / Big Data

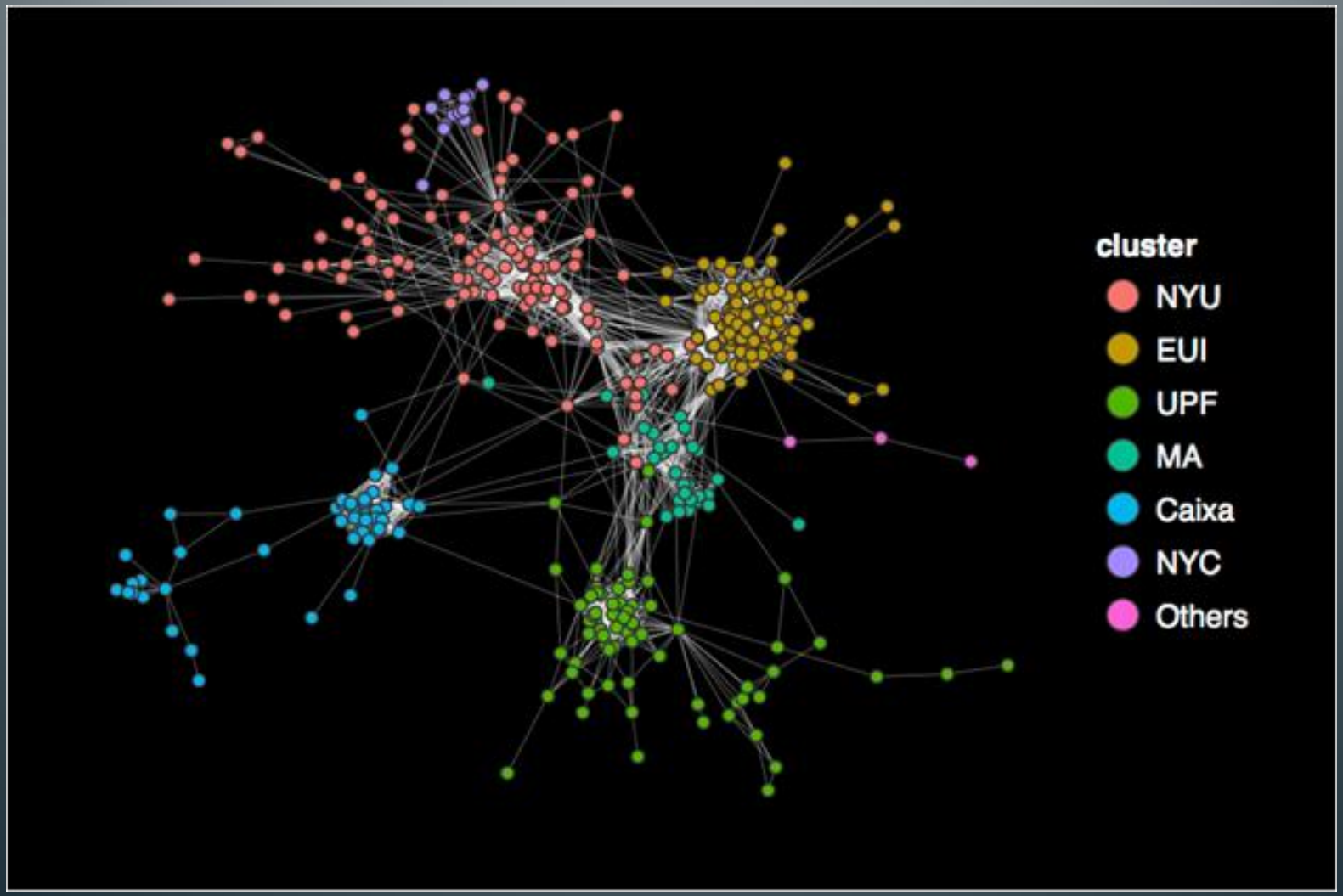
Все работы выполнялись с участием студентов кафедры КМАД

Разработка алгоритмов и программного обеспечения для классификации объектов в Интернете Заказчик «Samsung Electronics Ukraine Company» LTD (SEUC)

По заказу SEUC кафедрой по контракту были выполнены две прикладные НИР:

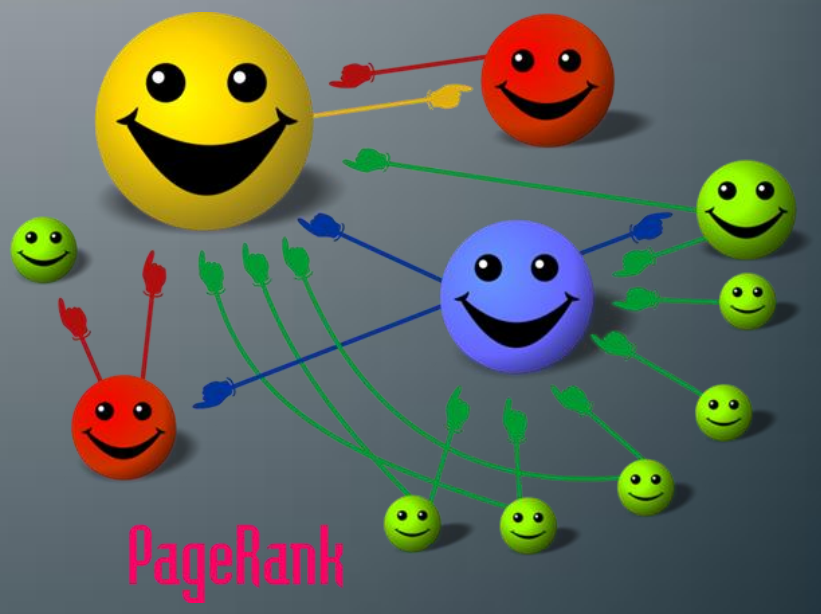
- "Автоматизированная (полуавтоматическая) кластеризация большого количества объектов по ограниченным текстовым описанием" / "Semi-supervised clusterization of large corpora of short texts".
- "Алгоритмы частичного машинного обучения» / "Semi-supervised machine learning algorithms “.

Кластеризация текстовых объектов

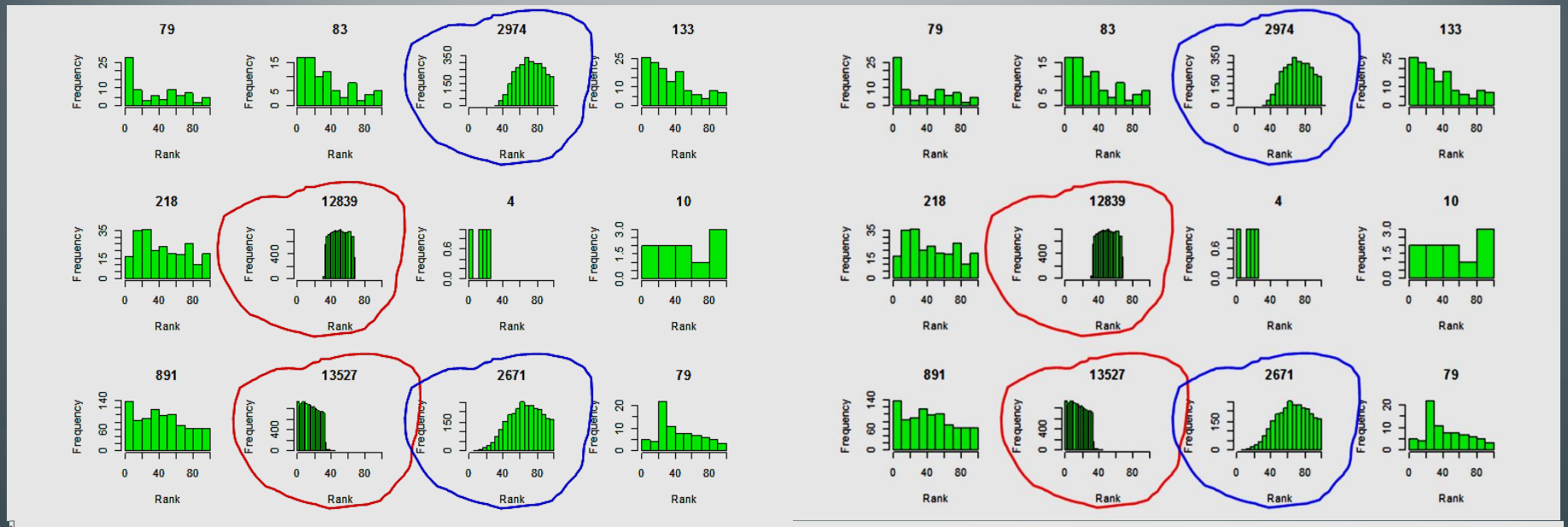


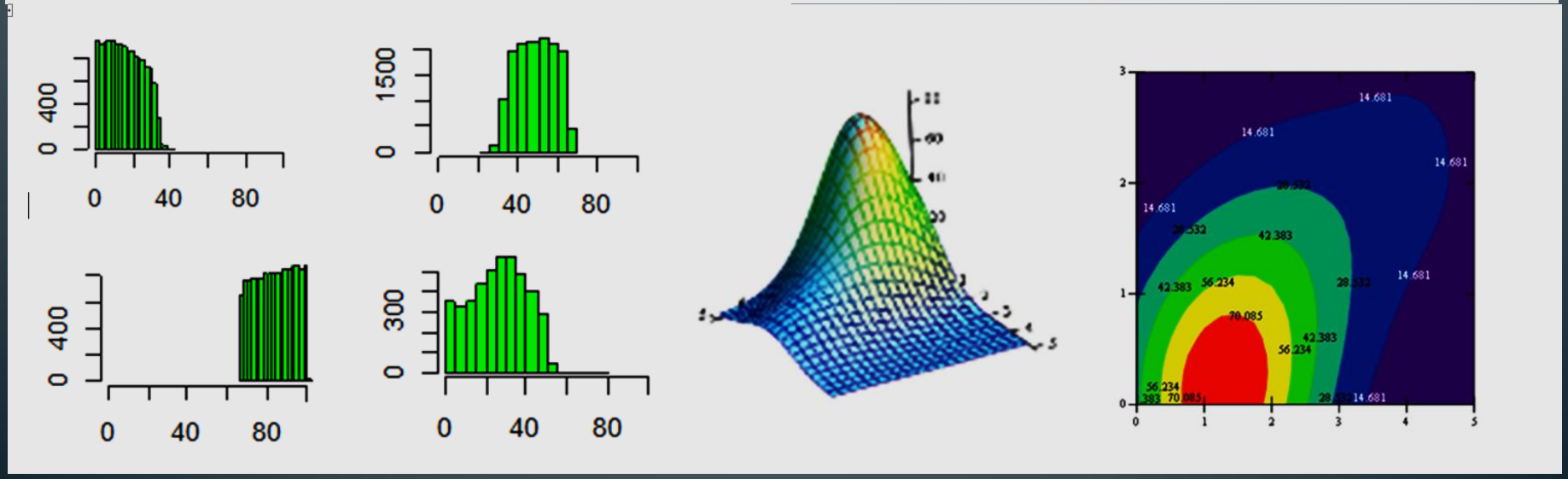
Разработка адаптивных моделей ранжирования сайтов в поисковых системах на основе методов «Ranking Learning»
Заказчик “Noosphere Ventures” & TrafficDNA.com

Разработана методика, алгоритмы и программное обеспечение для построения обучающихся моделей систем ранжирования сайтов на основе динамической кластеризации



ПОСТРОЕНИЕ ОБУЧАЮЩЕЙСЯ МОДЕЛИ РАНЖИРОВАНИЯ САЙТОВ НА ОСНОВЕ КЛАСТЕРИЗАЦИИ





Разработка алгоритмов и программного обеспечения для статистического анализа и прогнозирования финансовых рынков
Заказчики Math+.com & Tradespoon.com, USA

Разработаны методы и программное обеспечение для анализа и прогнозирования финансовых рынков

Создана WEB – базированная интеллектуальная система поддержки принятия решений при торгах на электронных финансовых биржах

ДАННЫЕ ON-LINE МОНИТОРИНГА ЭЛЕКТРОННЫХ ТОРГОВ НА ФИНАНСОВОМ РЫНКЕ

СТАТИСТИЧЕСКИЙ АНАЛИЗ ДАННЫХ ФИНАНСОВЫХ РЫНКОВ

The 8th IEEE International Conference on Intelligent Data Acquisition and Advanced Computing Systems: Technology and Applications
24-26 September 2015, Warsaw, Poland

Nonlinear Signal Reconstruction Based on Recursive Moving Window Kernel Method

Leonid Lyubchyk[1], Vladislav Kolbasin[1], Roman Shafeyev[1]
[1]National Technical University "Kharkiv Polytechnic Institute",
Frunze str. 21, Kharkov, 61002, Ukraine;
lyubchyk@kpi.kharkiv.edu

***Abstract*— Reconstruction problem for signals generated by discrete nonlinear dynamic system is considered via unified approach to recurrent kernel-based dynamic systems. In order to prevent the model complexity increasing under on-line identification, the reduced order model kernel method is proposed and proper recurrent Least-Square identification algorithms are designed along with conventional regularization technique. The recurrent version of Moving Window Kernel Method is also considered and suitable identification algorithm is developed, which has tracking properties and may be successfully used for on-line identification of nonlinear and nonstationary signal reconstruction.**

***Keywords*— kernel methods, machine learning, nonlinear signal reconstruction, system identification, support vector machine, recurrent least-squares.**

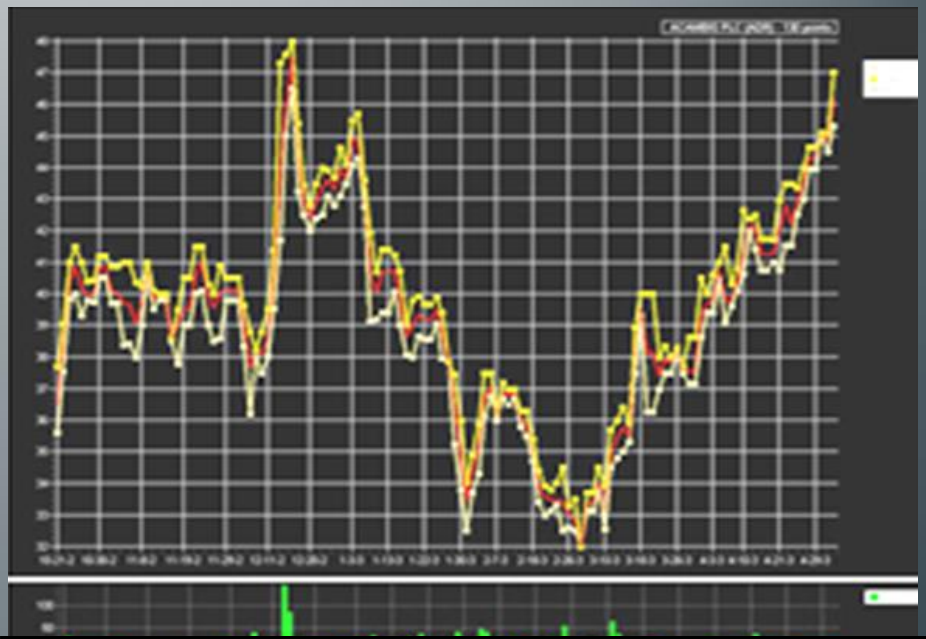
I. Introduction

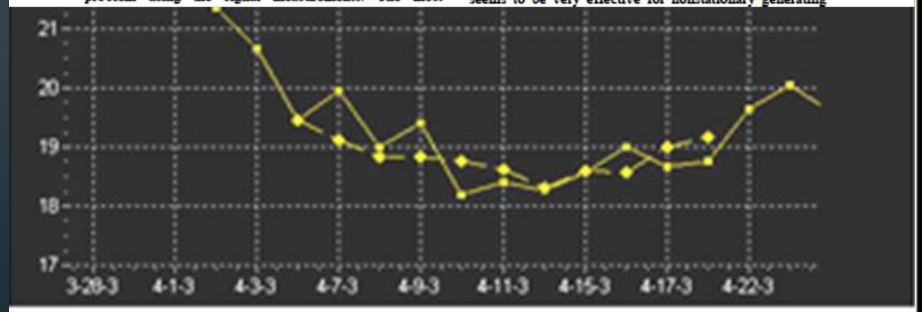
The problems of signal reconstruction from discrete samples are of a great importance in different applications, such as signal processing, automatic control, econometrics etc [1]. At that recovered signal is usually treated as an output of the discrete dynamical system with unknown mapping function that allows reducing the signal reconstruction to an equivalent system identification problem using the signal measurements. The most approaches, which produce nonlinear and nonparametric versions of conventional identification algorithms.

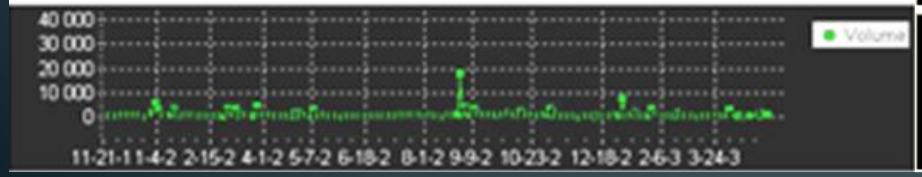
Kernel identification methods, based on the idea of input data implicit nonlinear transformation into high-dimensional (theoretically infinite) feature space, ensure the possibility of nonlinear model approximation. Using the Mercer's theorem, the feature vectors are chosen so that its scalar products in feature space are the positive definite kernel functions. At that the identified model may be represented in a nonparametric form as linear combination of kernel function, though the weighting coefficients (auxiliary variables) may be computed without making direct reference to feature vectors [7]. Such an idea is proved to be very effective for nonlinear system identification [8].

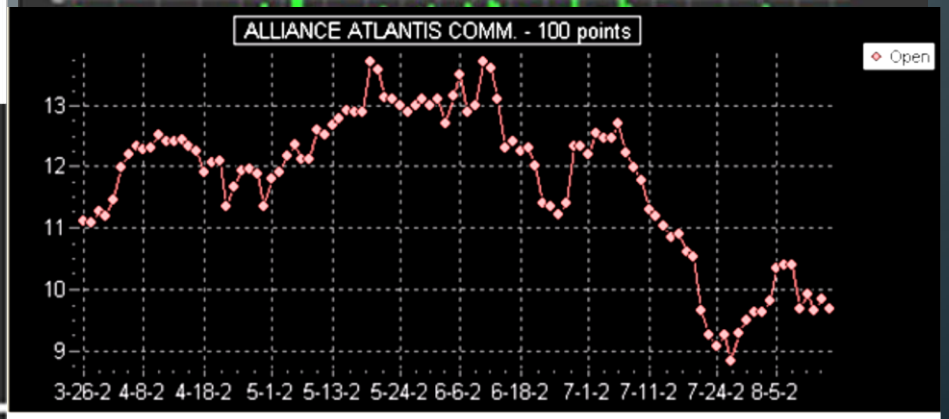
In the initial version of KM the dimension of auxiliary variables vector as well as estimated model complexity increases proportionally to the measurements sample length, which make such approach unsuitable for on-line signal processing. In order the model complexity restriction, it is desirable to use the auxiliary vector of fixed dimension along with recurrent version of identification algorithm [9].

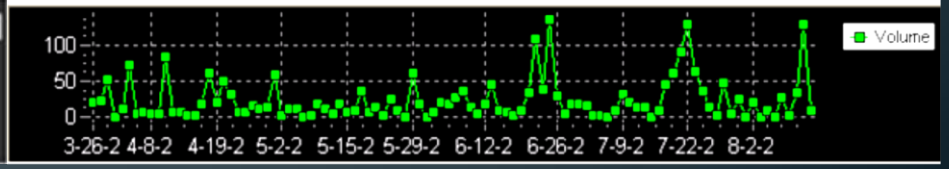
In [10] the SVM approach to Recursive Least-Square Kernel Method (RLSKM) has been considered, which seems to be very effective for nonstationary generating





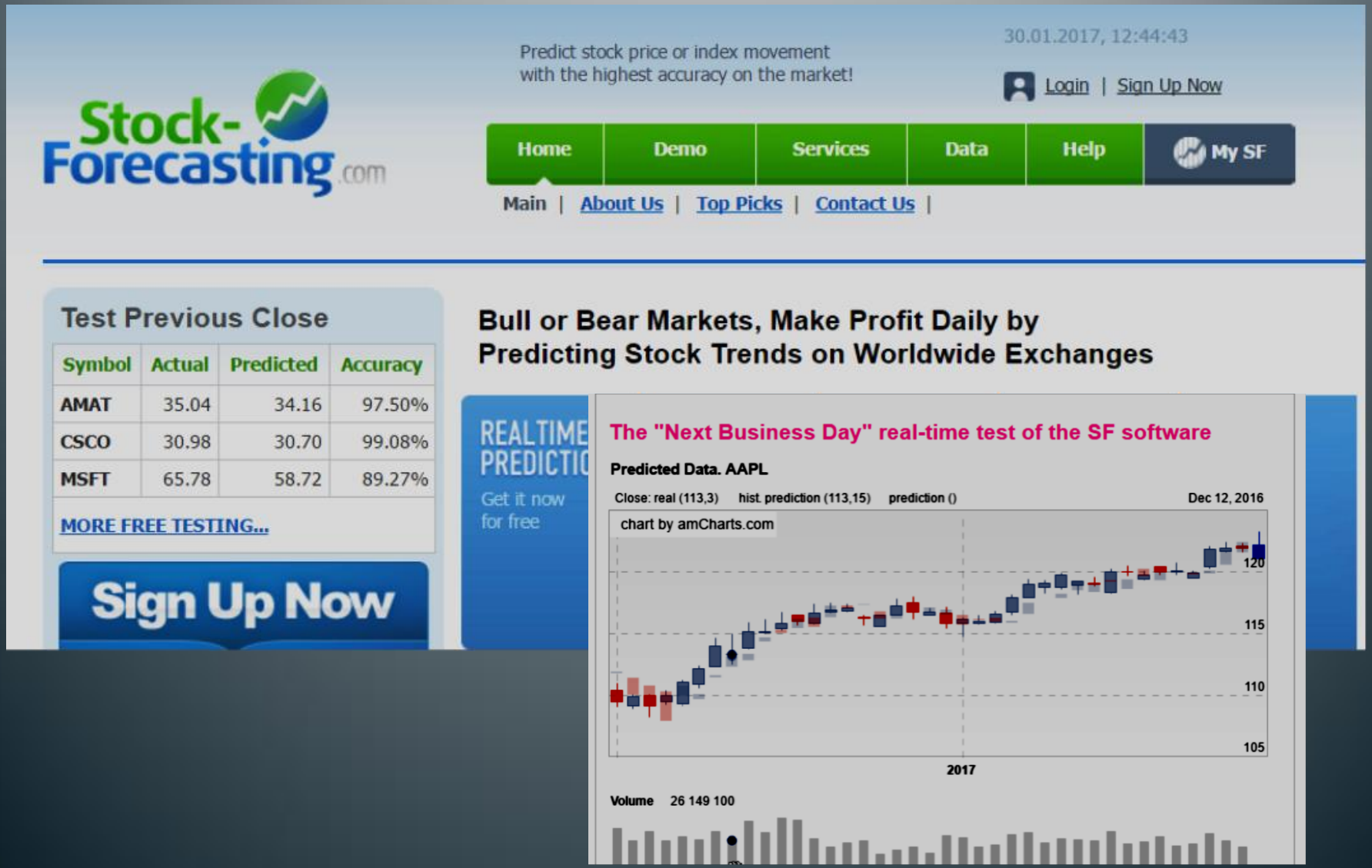






КОММЕРЧЕСКИЙ ПРОДУКТ

WEB-БАЗИРОВАННАЯ СИСТЕМА ПОДДЕРЖКИ ПРИНЯТИЯ РЕШЕНИЙ ПРИ ТОРГАХ НА ЭЛЕКТРОННОЙ ФИНАНСОВОЙ БИРЖЕ



Студенты кафедры КМАД дважды выходили в финал Чемпионата мира по программированию (International Programming Contest ACM-ICPC). Харбин, Китай

Certificate of Achievement

acm International Collegiate Programming Contest

IBM event sponsor

The 34th Annual ACM-ICPC
World Finals
Harbin 5 February 2010

awarded to

National Technical University
Kharkiv Polytechnic Institute

Daniil Bondarev
Stanislav Cherviakov
Dmytro Dzhulgakov
Leonid Lyubchyk, Coach

Honorable Mention

from among 103 teams chosen from a field of 7,319 teams from 1,931 universities in 82 countries competing on six continents.

William B. Poucher, Ph. D.
ICPC Executive Director

Desen Yang, Professor & Vice President
Harbin Engineering University

Ypsilon Pi Epsilon

Honor Society for the Computing & Information Disciplines
in recognition of your team's participation at the

The 2010 ACM International Collegiate Programming Contest
Sponsored by IBM

this

CERTIFICATE OF ACHIEVEMENT

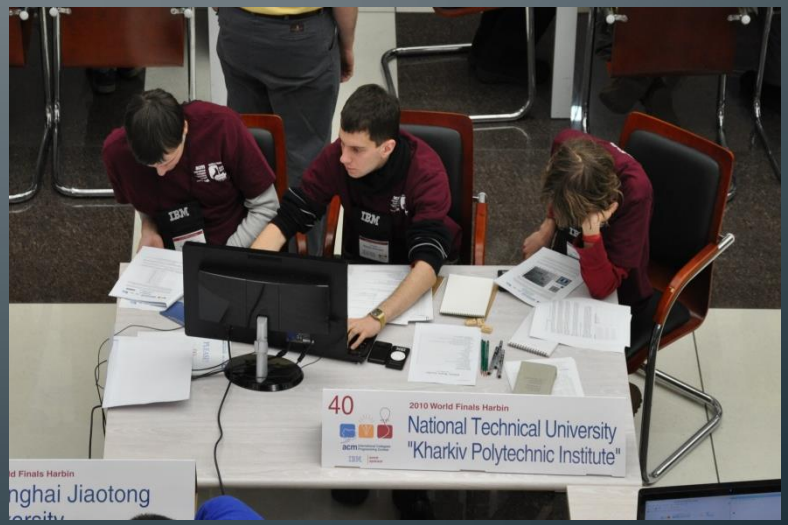
is presented to

National Technical University "Kharkiv Polytechnic Institute"

Team Members

Daniil Bondarev, Stanislav Cherviakov, & Dmytro Dzhulgakov
Leonid Lyubchyk, Coach

Harbin, China
February 1-6, 2010



Финал Чемпионата мира по программированию (International Programming Contest ACM-ICPC), Варшава, Польша.

acm icpc

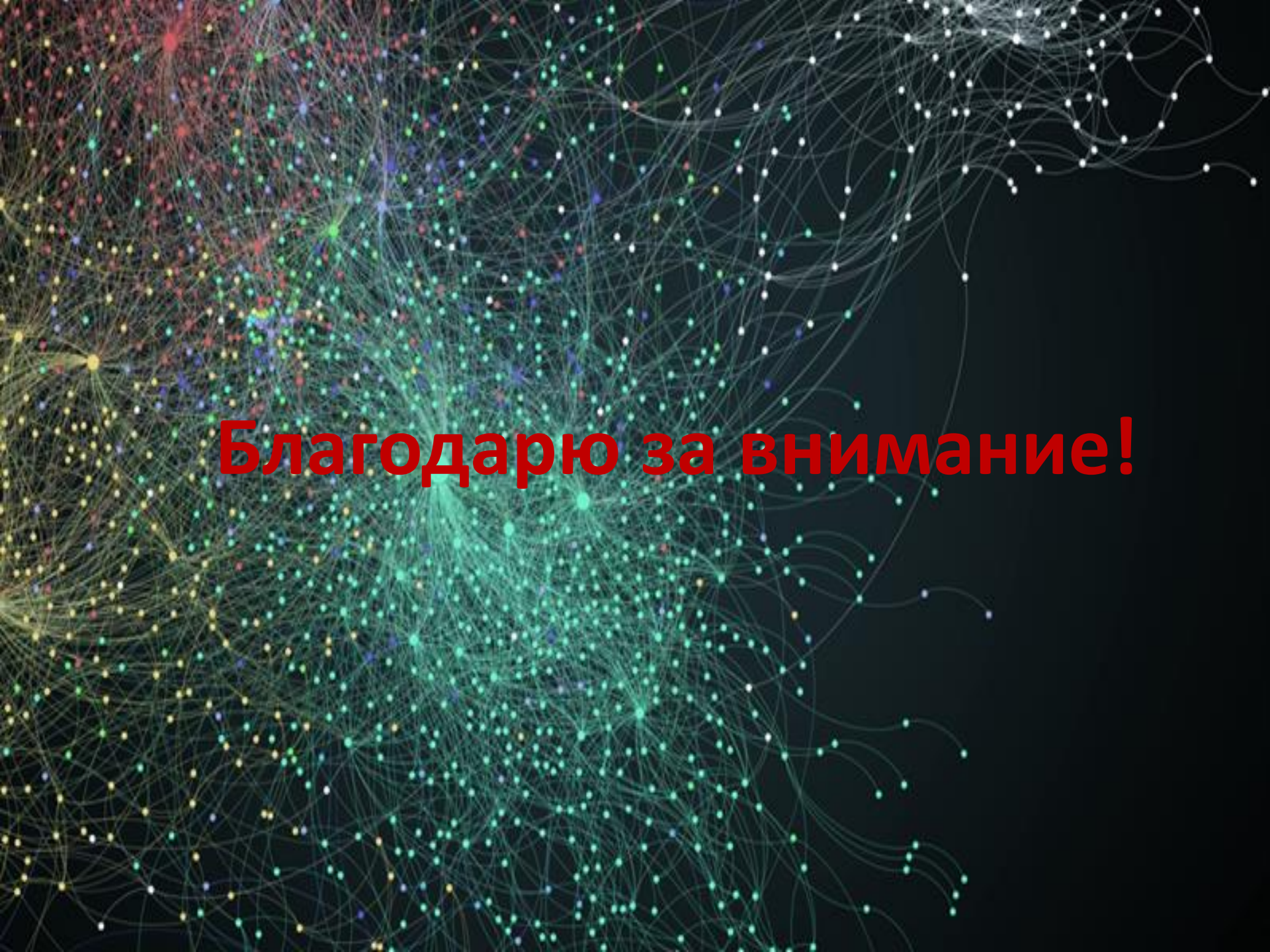
IBM event sponsor

international collegiate programming contest

Ждем Вас на кафедре КМАД НТУ «ХПИ»!

Лучшие выпускники кафедры

- Джулгаков Дмитрий – Software Engineer in Facebook, starting 2012.
- Полосухин Илья – Software developer and Data Scientist in Salford Systems , 2008 – 2014, Software Engineer in Google, starting 2014 .
- Бичурин Андрей - Software Engineer Intern in Google, 2013, Research Engineer in Samsung, 2013-2014, Software Engineer in Google, starting 2015
- Лисичкин Владислав – Software Engineer Intern in Google, 2015, Research Engineer in Samsung, 2013 – 2014, Software Engineer in Google starting 2015.
- Бондарев Даниил - SDE2 (Software Developer Engineer) in Amazon Web Services
- Никульченко Артем – стажировался в IBM Zurich Research Lab, сейчас заместитель директора Cloud Works



Благодарю за внимание!