

Современные тренды в области Data Science. По материалам конференции Data Fest 5

28 апреля

Data Fest⁵

Москва
FLACON

Контаев Илья

mail: nerdikon@gmail.com

tg: @ikont

Как это было

Крупнейшая конференция по Data Science в России и СНГ

Место действия: дизайн-завод FLACON

Data Fest 1(2015): ~400 человек

Data Fest 5(2018): ~5000 человек

Около сотни докладов в 2018 году



Виды докладов

— — —

- 📄 **Передовая ML, DL, модели и как они работают**
- 📄 Инструменты и инженерная начинка решений
- 📄 Как DS и ML приносят пользу бизнесу
- 📄 Интерактивные секции от сообщества

	Main stage	Prod stage	Space	Lumiere hall 1	Lumiere hall 2
10:30					
11:00					
11:30	11:00 - 12:45 Russian ML keynotes, part 1	11:00 - 12:45 Big Data	11:00 - 12:45 PyData	11:00 - 12:45 ML training	11:00 - 12:45 Pitch session
12:00					
12:30					
13:00	13:00 - 14:30 Russian ML keynotes, part 2	13:00 - 14:30 Sys ML, part 1	13:00 - 14:30 BigARTM workshop	13:00 - 14:30 Recsys workshop	13:00 - 14:30 DeepTensor workshop
13:30					
14:00					
14:30					
15:00					
15:30	15:30 - 17:30 International ML keynotes	15:30 - 17:30 Sys ML, part 2	15:30 - 17:30 DeepBayes workshop	15:30 - 17:30 Deep Computer Vision workshop	15:30 - 17:30 iPavlov workshop
16:00					
16:30					
17:00					
17:30					
18:00	17:45 - 19:15 ML in business	17:45 - 19:15 Fit ML	17:45 - 19:15 True story case-club	17:45 - 19:15 NLP in industry	17:45 - 19:15 WiML
18:30					
19:00					
19:30	19:30 - 21:00 ML puzzler 🧩	19:30 - 21:00 ML failconf	19:30 - 21:00 DS in E-sports	19:30 - 21:00 Sigbovik workshop	19:30 - 21:00 Unconference
20:00					
20:30					
21:00					

Keynote by Константин Воронцов

— — —

[Слайды](#) [Видео](#)

- [Отчет белого дома США по ИИ](#), 2016 год
- Особенности решения задач ML на реальных данных
- Взаимодействие бизнеса, науки и образования
- Соревнования на основе открытых данных
- Платформы для онлайн обучения моделей
- Вытеснит ли DL весь ML?(спойлер, нет)

Embeddings for Graph Classification by Евгений Бурнаев

[Слайды](#) [Видео](#)

Различные подходы:

- Статистики(#вершин/ребер/подграфов, распределения)
- Graph Kernels
- Anonymous Walk Embeddings
- graph2vec(graph is doc, anonymous walks are words)

RecSys by Яндекс.Дзен

[Слайды](#) [Видео](#) (keynote) [Видео](#) (workshop)

- *Евгений Соколов, ФКН ВШЭ*: Ранжирование новых форматов в Дзене
- *Борис Шарчилов, Яндекс.Дзен*: Поиск влиятельных примеров для градиентного бустинга
- *Дмитрий Ушанов, Яндекс.Дзен*: Обзор нейросетевых рекомендательных систем

RecSys by Яндекс.Дзен

— — —

- Качество vs Интерпретируемость моделей
- Поиск влиятельных примеров в моделях NN и GB
- Современные модели нейросетей для RecSys:
 - DSSM
 - Neural Collaborative Filtering
 - Cross-domain Content-boosted CF
- Примеры использования в Дзене

Deep Tensor by Skoltech, Scientific Computing Group

— — —

[Слайды](#) [Видео](#)

- *Евгений Фролов*: Моделирование предпочтений пользователей с помощью тензорных методов
- *Валентин Хрульков*: Атаки на нейронные сети
- *Александр Новиков*: Тензоры и рекуррентные нейронные сети
- *Александр Катруца*: Deep Multigrid
- *Алексей Гринчук*: Сжатие нейронных сетей при помощи тензорных методов

Deep Tensor by Skoltech, Scientific Computing Group

— — —

- Тензорная модель Collaborative Full Feedback для RecSys
- Типы adversarial атак на нейронные сети
- Tensor Train RNNs. Теоретическое доказательство эффективности RNN
- [T3F](#)(Tensor Train on TensorFlow)
- Использование идей из DL для решения СЛАУ(хардкор)
- Tensor Train разложение как способ сжатия нейросетей

BigARTM и тематическое моделирование

— — —

[Слайды](#) [Видео](#)

- *Константин Воронцов* : Тематическое моделирование в BigARTM: новые возможности
- *Анастасия Янина (МФТИ)*: Иерархические тематические модели для разведочного поиска
- *Николай Скачков (МГУ)*: Тематическая сегментация текста
- *Илья Жариков (МФТИ)*: Гиперграфовые тематические модели транзакционных данных
- *Геннадий Штех*: Как BigARTM помогает составлять плей-листы
- *Арина Агеева* : Как BigARTM помогает отбирать релевантных кандидатов по резюме

BigARTM и тематическое моделирование

- Обзор задач в тематическом моделировании
- Иерархическое тематическое моделирование
- Разведочный поиск
- Тематическая модель гиперграфа. Анализ транзакционных данных
- Тематические эмбединги
- Приложения: создание музыкального плейлиста, подбор персонала

DeepPavlov by МФТИ

— — —

[Слайды](#) [Видео](#) (keynote) [Видео](#) (workshop)

- *Михаил Бурцев*, Разговорный интеллект сегодня(keynote)
- *Валентин Малых*, Что внутри у диалоговой системы?
- *Александр Селиверстов*, Open source для чат ботов и диалоговых систем
- *Михаил Архипов*, Распознавание именованных сущностей
- *Диляра Баймурзина*, Классификация текстов
- *Мария Вихрева*, Гибридный диалоговый бот
- *Алексей Лымарь*, Автоматическое исправление опечаток
- *Леонид Пугачев*, Использование ранжирования
- *Юрий Куратов*, Поиск ответа в тексте
- *Максим Кретов*, Overview of research ideas

DeepPavlov by МФТИ

— — —

- State of the art диалоговых систем
- Общее и внутреннее устройство библиотеки DeepPavlov
- Модули: NER, Intent classification, Dialogue Manager, NLG
- Семантическое ранжирование
- Поиск ответа на вопрос в тексте. Модель [R-Net](#)

Deep Bayes by HSE, Bayesian Methods Research Group

— — —

[Слайды](#) [Видео](#) (keynote) [Видео](#) (workshop)

- *Гадецкий Артём, Higher School of Economics: Обусловленные генераторы определений слов*
- *Полыковский Даниил, Insilico Medicine, Moscow State University: Применение глубокого обучения для поиска лекарств*
- *Молчанов Дмитрий, Samsung AI Research: Дисперсионные сети*
- *Ашуха Арсений, University of Amsterdam, Samsung AI Research: Вариационный дропаут для разреживания нейронных сетей*
- *Чиркова Надежда, Higher School of Economics, Samsung-HSE laboratory: Байесовское разреживание рекуррентных нейронных сетей*

Deep Bayes by HSE, Bayesian Methods Research Group

— — —

- *Атанов Андрей, Higher School of Economics, Samsung-HSE laboratory: Стохастический батчнорм для оценки неопределенности*
- *Иванов Олег, Lomonosov Moscow State University: Нейробайесовское восстановление пропущенных данных*
- *Новиков Александр, Higher School of Economics, Samsung-HSE laboratory: Тензорные разложения и их использование в машинном обучении*
- *Кузнецов Максим, HSE/Skoltech, Insilico Medicine: Тензорный взгляд на дискретные вариационные автоэнкодеры*
- *Кропотов Дмитрий, Lomonosov Moscow State University: Методы оптимизации и сэмплирования для нейросетей через K-FAC*

Deep Bayes by HSE, Bayesian Methods Research Group

— — —

- Генерация определения по слову как способ интерпретации
- Дисперсия вместо весов в нейросетях
- Variational Dropout для уменьшения сложности моделей
- Восстановление пропущенных данных с помощью Variational Autoencoder
- Tensor Train разложение как способ сжатия нейросетей(again?)

Спасибо за внимание!



The Great Wall of China
Harold Lloyd Jones



The Yellow Sky
Paul Gauguin

Яндекс

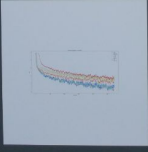


Loss Function Art

В сентябре 2017 года мы запустили проект Loss Function Art — проект, посвященный искусству, созданному с помощью нейронных сетей. На сегодняшний день проект уже достиг отметки в 100 тысяч произведений искусства.

Каждое искусство создано с помощью нейронных сетей, на основе данных из публичных баз данных. Каждое искусство создано с помощью нейронных сетей, на основе данных из публичных баз данных.

Сейчас проект уже набрал популярность. С каждым годом количество произведений искусства, созданного с помощью нейронных сетей, увеличивается. В будущем мы планируем расширить проект и добавить новые функции.



The Yellow Sky
Paul Gauguin



The Yellow Sky
Paul Gauguin