



Please find more info about the sessions presented by Speakers below:

DESCRIPTIONS

Chenglong Chen - Deep Learning Formula for NLP Applications

This presentation will talk about the general pipeline for deep learning with text. It can be summarised as a four-step strategy: embed, encode, attend and project. We will show through examples how they're put together to solve various NLP tasks such as text classification, semantic matching, sequence labelling and more. Approaches to improve the performance of NLP model will also be discussed.

Qingchen Wang - Machine Learning for optimization problems

This session focuses on using machine learning for solving optimization problems. There are many optimization problems that are traditionally solved either by deriving complex mathematical approximations or by programming simulations. However, traditional solutions are often limited by computational complexity. An example of such a problem is the general workforce planning problem for call centers or hospitals, where the planner tries to find the best set of agents and schedules to ensure high quality of service but at minimal costs. For these problems, I show how machine learning can be used to solve this problem. During this session, I will first describe the problem in detail, then show how I machine learning in combination with simulations and feature engineering can be used to find better solutions to these optimization problems than traditional approaches.

Qingchen Wang - Kaggle step-by-step 2

This session is a tutorial on how to approach Kaggle competitions. I will show the audience some basic steps of setting up a good validation framework, explain how the different evaluation functions work, go through some feature engineering steps/tricks, and given an overview of the value of Kaggle's public kernels. The content of the presentation will be within the context of real competitions and I will give examples based on actual competition data.

TOPICS:

- Data cleaning
- Feature engineering: numeric, categorical features, datetime and text, brief overview
- Main ML algorithms
- Metrics and objective function

Zhifeng Gao - Details and implementation in GDBTs

This session will be about underlying and interesting details in gbdt algorithms, like xgboost, lightgbm and catboost. I will share some common and different in those popular GBDT models.

Zhifeng Gao - Feature Engineering for Tabular Data

Basic and important tips in feature engineering, specifically, some experience from Speaker's perspective in feature engineering about derive feature, feature encoding, target encoding and feature decomposition.

Mikhail Trofimov - Common obstacles in competitions and how to avoid them

This session is a novice-friendly tutorial and about solution steps and common mistakes in competitions. I will also cover the topic of leak types and validation scheme. Mostly tips, tricks and best practices.

Michael Yeh - Workshop for experts: Fine-tune BERT in Pytorch with tricks.

I'll go through about the automatic mixed precision (AMP) concept: how it speeds up the training, setup in PyTorch. Then use PyTorch to build BERTXLNET, load pre-trained weight, run training codes. With extra techniques like minibatch accumulation to stabilize the fine-tuning, tensor-trimming to accelerate the inference. We will use the data in the latest Jigsaw competition.

Michael Yeh - Presentation for novices: Journey of an Addicted Kagglers

Inspirational talk about getting into Data Science and Kaggle field, how to start and what can you learn from Kaggle.

Yuanhao Wu - Advanced Language Models for NLP Problems (expert track)

This session focuses on latest language models such as BERT, Transformer-XL, GPT and XLNet. These models outperform traditional models such as LSTM, GRU with a significant margin, and they have set the new baselines for many NLP problems. I will introduce these models and give some examples of how to use these models in different NLP problems.

Yuanhao Wu - BERT for Kaggle Competitions (novice track)

In this session, I will introduce the BERT model. I will talk about the theory, finetune tricks and my experience in the Jigsaw Unintended Bias in Toxicity Classification Competition.

Pavel Ostyakov - Expert track. Kaggle Pixel Pro - building your own AI Camera.

These days, all modern smartphones post-process photos taken from their cameras. Using deep learning approaches, they increase resolution, remove noise, enhance colors, and make night photos beautiful. In this section, we will talk about computational photography and how to use computer vision to improve the quality of images. We will consider popular techniques and build our pipeline for images processing.

Pavel Ostyakov - Novice track. Computer vision for dummies.

In this section, we will learn how to use convolutional neural networks on the example of the image denoising task. We will see how to prepare data, design a neural network's architecture, and train it using PyTorch library. We will consider popular approaches for improving quality and speeding up experiments.

Dmitry Larko

Beginner track - Time series 101.

In his presentation Dmitry will share a common approaches for time-series forecasting and model validation which helps you to start you own forecasting in this space.

Dmitry Larko

Expert track - Deep Learning for time series.

The main topic of this presentation is the application of Deep Learning to time-series forecasting and classification. Dmitry will share common approaches to data preparation, architecture design and validation of neural networks used for time-series classification or forecasting.

Jun Lan

Tricks for image classification

I will show tricks for image classification which I have learned from past kaggle competitions, such as mixup, knowledge distillation, pseudo label etc.

Yauhen Babakhin

Topic: Common errors in Machine Learning pipelines

Description: In this workshop we will cover the most common errors one could make while developing the Machine Learning / Deep Learning pipeline either for a real project or a Kaggle competition. We will start with an initial pipeline and try to locate and correct all the possible errors / bugs it contains.

Kazuki Onodera

Essential techniques for tabular competition with case study

He will describe some of his techniques for tabular competition with case studies. You will understand what you should work on first.

Tao Shen - Title: Winning Kaggle competitions with deep learning skills

This presentation will talk about my experience in winning Kaggle competitions with deep learning skills. In this session, I will first Introduce the characteristics of deep learning competitions in CV, NLP and ASR field, then show the general modeling method of deep learning, give examples on some finished competitions and sum up some important notes/tricks in winning such competitions.