ML in PL Conference

ML in PL - Contributed Talks

This is an open list of speakers that were chosen through the ML in PL Call for Contributions along with their presentation titles and abstracts. The ordering is arbitrary.

The full conference program can be found here.

Adam Dendek

(AGH University of Science and Technology)

Machine Learning in High Energy Physics

High Energy Physic is a branch of physics that studies the nature of the particles that constitute matter and radiation. It study the most fundamental building blocks of nature by utilizing the accelerators producing energies not seen since the Big Bang. The LHC accelerator and experiments program has lead to significant discoveries such as the first observation of the Higgs Boson.

On one side precision measurements require precise calibrations, on the other searches for rare SM processes and for beyond-SM processes require sifting to a huge amount of data in a search of a very small signal. With these objectives, Machine Learning techniques play a crucial role.

The presentation will cover the research conducted by my home group at AGH. I will focus on such Machine Learning applications as track reconstruction, calibration anomaly detection, detector response generation, and others.

Adam Gonczarek & Szymon Zaręba *(Alphamoon)*

Few-shot, semi-supervised and transfer learning for NLP, or, how to perform well on NLP tasks when lacking labeled data

Methods for intelligent text analysis are becoming highly desirable in various industries including banking, insurance, law, and healthcare. Recently developed models such as sequence-to-sequence or the Transformer allow for building powerful methods for many real-life NLP problems in an end-to-end deep learning fashion. However, having enough labeled examples to train them in a supervised manner usually remains a challenge. In this talk we will show how to deal with that problem by taking advantage of several techniques, including few-shot learning with Relational Networks, semi-supervised learning with Pi-Models and entropy-based losses, and transfer learning with BERT. We will show how to adopt them properly for two problems: document classification and named entity recognition. Finally, we will highlight several challenges that we faced when handling Polish and English languages.

Adam Zadrożny *(IUS.AI / NCBJ)*

Towards Understanding Polish Court Verdicts

We have developed an artificial intelligence (AI) powered system for understanding legal texts of Polish court verdicts. Using novel AI methods and our state-of-the-art natural language processing tools we were able to construct a system that was able to search for most relevant court verdicts for a given legal question. The texts of court verdicts were extracted from public databases. Our system was tested on cases prepared for a tax advisor exam and solved multiple of them. Since summer 2019, the system in the beta version has been used by a few legal companies.

Adrian Łańcucki *(University of Wroclaw)*

Unsupervised Neural Segmentation and Clustering for Unit Discovery in Sequential Data

We study the problem of unsupervised segmentation and clustering of handwritten lines with applications to character discovery. To this end, we propose a constrained variant of Vector Quantized Variational Autoencoder (VQ-VAE) which produces a discrete and piecewise-constant encoding of the data. We show that the constrained quantization task is dual to a Markovian dynamics prior placed on the latent codes. Such view facilitates a probabilistic interpretation of the constraints and allows efficient inference. We demonstrate the effectiveness of the proposed method in the context of unsupervised handwriting character discovery in 17th-century scanned manuscripts. Lastly, we discuss and address some of the known difficulties in training discrete VQ-VAE representations, which stem from improper initialization and scale mismatch between encoded representations and learned latent codes. This work has been done during the 2019 JSALT Summer Workshop https://www.clsp.jhu.edu/workshops/19-workshop/distant-supervision-for-representation-learning-in-speech-and-handwriting/ .

Aleksander Obuchowski

(SentiOne / Politechnika Gdańska)

Automated answer prediction in customer service FAQ conversations based on question-question similarity and answer clustering

Customer service has been gradually transforming since the social media boom in 2010. With the increase in communication channels, e.g. chat, the number of messages requiring action has been growing. Thus automation has become inevitable in this area with solutions accelerating the reaction time of agents. On the other hand, this shift in communication has made conversational data available for analysis. Questions asked by customers tend to be repetitive (of an FAQ-nature) and receive a simple, standardized answer, written by customer service agents. Such data structure is similar to corpora of the FAQ-related NLP tasks, e.g. community question answering (see the tasks of SemEval 2017), and allows for the design of answer retrieval or answer prediction architectures. We introduce a system of automated response suggestions in the Polish language, offering answers to FAQs based on historical conversations between users and customer service in the banking area. Our solution focuses on finding similar questions that already have an answer to them, thus using the concept of answer retrieval via question to question similarity [1]. For that approach we propose a novel way of building a question encoder based on a neural-network, pre-trained in unsupervised fashion by clustering the answers and predicting them based on questions. The architecture has been evaluated on a Polish customer service corpus in the banking domain. The solution has proven to be better than the current state-of-the-art approaches while achieving 86% top-3 accuracy on the response prediction task compared with 68% of the Google encoder. This method has proven to provide state-of-the art results, reducing top-3 error rate by 44%. The solution is applicable to other domains and languages, if similar corpora are provided for encoder pre-training.

1. Nakov P. et al. 2017. Semeval-2017 task 3: Community Question Answering. In Proceedings of the 11th Int. Workshop on Semantic Evaluation, pp. 27-48. ACL

Anna Dawid *(University of Warsaw)*

Can a learning machine teach us quantum physics?

Machine learning affects everyday life in manifold ways, like character and voice recognition software, fingerprint identification, e-mail spam filtering, self-driving cars, and many others. Its impact is also prominent in science, as neural networks have been harnessed to solve problems of quantum chemistry, astrophysics, material science, and biology, i.e. fields characterised by a large amount of data. Quantum many-body physics struggles with a similar dimensionality curse, and unsurprisingly, a booming research worldwide is devoted to machine learning application in quantum problems, especially those in phase classification. However, had this new approach provided any new results concerning phase classification? So far it has enabled only the recovery of known phase diagrams or the location of phase transitions with qualitative agreement with more conventional approaches (but at much lower computational cost). Can physicists be at least sure that the machine learns anything related to the so-called order parameter which in the physics framework is used to describe and define phase transition? Not really.

We show that interpretability methods, well-known to machine learning community, like influence functions or heat maps, can be used to address this question in the context of phase classification by convolutional neural networks (CNNs). Their indications can be threefold: firstly, they can show that CNNs indeed learn an order parameter, secondly, that CNNs learn some noise, background, other information unconsciously provided by a researcher, or, what would be especially exciting, they learn something relevant for physics, different than order parameter, but not noticed before. In such a way, they indeed can teach us quantum physics.

Antoni Sobkowicz *(Ośrodek Przetwarzania Informacji - Państwowy Instytut Badawczy)*

Voting haters - trying to predict election results using online comments

Political news are a major part of many information websites and are often emotionally commented on by a huge number of users. Using such data - millions of comments - we searched for an automated way to find coherent groups of users that produce content that is likely to evoke negative emotions towards political parties. Combining time patterns with machine learning methods we build a toolkit to infer an opinion poll using web public data. In this talk we'll delve into how what we achieved, how we did it and what can be done in the future.

Barbara Rychalska

(Findwise / Warsaw University of Technology)

Neural Machine Translation: achievements, challenges and the way forward

The talk will present insights from cooperation between Warsaw University of Technology and Nanyang Technological University Singapore that I have been part of during summer 2019. I want to speak about the current state of the art in Neural Machine Translation: what is possible, what is not possible, what are the common misconceptions about applying NMT to real world problems, and the many challenges connected with evaluating these models. I will tell about adversarial examples that fool NMT models (and in fact, all models!), which can originate by mistake or by purpose - via actions of a malignant attacker. I will approach the problem from 2 perspectives: sentence-level translation and document-level translation and show examples of the machines' level of understanding context. The main goal is to present the promising results that originated from our joint work at WUT and NTU, and promote participance in this project within MI2Datalab at Warsaw University of Technology.

Bartłomiej Borzyszkowski & Michał Bień

(CERN Openlab)

AI at CERN: Neuromorphic Computing in High Energy Physics

Together, we will introduce the program at CERN OpenLab and make an overview of the European Organization for Nuclear Research. After this brief introduction, we will discuss our separate projects:

Neuromorphic Computing in High Energy Physics - Investigating the potential of Spiking Neural Networks deployed on neuromorphic chips (Intel Loihi processor) as a technological solution to increase the precision of the upgraded CMS detector for High-Luminosity LHC, the next stage of the LHC upgrade (scheduled to start in 2025);

Big Data and Machine Learning in the Cloud - Discovering new ways to use computing resources in hybrid cloud environment to address CERN machine learning computational capacity needs.

Cezary Olszowiec *(Comarch)*

On the solution concepts to the Reinforcement Learning problems

In the recent years we have been observing rapidly growing interest in the modelling of multi-agent systems, in particular when it comes to the applications of methods arising from Deep Learning and Game Theory. In this talk I will compare classical and novel solution concepts to the multi-agent systems, pros and cons of each of these as well related open problems.

I will give a brief introduction to mathematical formulation of a Reinforcement Learning problem in model-free and model-based framework with particular emphasis on the ideas behind the learning algorithms used in applications. I will introduce the notion of a Nash Equilibrium and other invariant objects as the potential solution concepts to the multi-agent systems. I will discuss their utility depending on the form of a considered system and the problem of finding them in practice. Lastly I will discuss the novel concepts introduced in learning algorithms in the recent years.

David Haber *(Daedalean AI)*

Opportunities and Challenges when Building AI for Autonomous Flight

Deep neural networks have demonstrated impressive performances on visual recognition tasks relevant to the operation of autonomous drones or personal electric air-taxis. For this reason, their application to visual problems, including object detection and image segmentation, is promising, and even necessary, for autonomous flight.

The downside of the increased model performance is higher complexity, which poses challenges to topics, such as, interpretability, explainability, and eventually certification of safety-critical aviation applications. How do you convince the regulators (and ultimately the public) that your model is robust to adversarial attacks? How do you prove that your training and testing datasets are exhaustive? How do you test edge cases when your input space is infinite and any mistake is potentially fatal?

Over the last few months, we have partnered with EASA (European Union Aviation Safety Agency) to explore how existing regulations around safety-critical applications can be adapted to encompass modern machine-learning techniques. In this talk, we will visit the different stages of a typical machine-learning pipeline to discuss design choices for neural network architectures, desirable properties for training and test datasets, model generalizability, and how to protect ourselves against adversarial attacks. Finally, we will consider the opportunities, challenges and learnings that may apply more generally when building AI for safety-critical applications in the future.

Dominik Belter *(Poznan University of Technology)*

Robots Understanding Human Environment - deep learning in mobile robotics

Autonomous robots should detect objects in the environment and understand their meaning to operate without human supervision. Recently, great progress was made in the field of CNN-based object detection and pose estimation. These methods allow detecting and grasping the objects by a mobile manipulating robot. However, the robot should be also capable to interact with articulated objects like doors to performs manipulation tasks and move freely in the indoor environment. We propose a system for the detection and estimation of kinematic structures observed by a mobile manipulating robot equipped with the RGB-D camera. We also show how to enhance the robot's perception system and generate images of rigid objects from various viewpoints. Finally, we show the application of neural networks in legged robotics to evaluate potential footholds for the quadruped ANYmal robot.

Jacek Cyranka *(esportsLABgg / University of Warsaw)*

Towards Reliable Reinforcement Learning and AI

I will present my research statement on reliable reinforcement learning pursued in my research group at University of Warsaw. The applications that we have in mind is the continuous optimal control of robotic agents (robotic arms, walking humanoids etc..) The techniques we are base upon involve some classical applied math including ordinary differential equations and research on verification of cyber-physical systems.

Kalina Jasinska, Piotr Wójcik *(Allegro.pl)*

Similarity learning for zero-shot product matching

In standard supervised classification problem setting, instances of all possible target classes are observed during the training phase. This is not necessarily so in all classification problems. In real life, one frequently needs to recognize an instance of a class based only on the description of the class, without prior observation of any instances. Such a task is known as zero shot classification and can be performed by humans quite well. However, standard supervised classification methods are not suited for solving it.

In the talk we will introduce the zero shot classification setting and discuss possible solutions based on deep metric learning. Building on our experience from work on product matching in Allegro, we will give you hands on insights on how to create production-grade zero-shot classification models.

Kira Kempińska

(Alphamoon / University College London)

Modelling urban networks using Variational Autoencoder

A long-standing question for urban and regional planners pertains to the ability to describe urban patterns quantitatively. Cities' transport infrastructure, particularly street networks, provides an invaluable source of information about the urban patterns generated by peoples' movements and their interactions. With the increasing availability of street network datasets and the advancements in deep learning methods, we are presented with an unprecedented opportunity to push the frontiers of urban modelling towards more data-driven and accurate models of urban forms.

In this study, we present our initial work on applying deep generative models to urban street network data to create spatially explicit urban models. We based our work on Variational Autoencoders (VAEs) which are deep generative models that have recently gained their popularity due to the ability to generate realistic images. Initial results show that VAEs are capable of capturing key high-level urban network metrics using low-dimensional vectors and generating new urban forms of complexity matching the cities captured in the street network data.

Krzysztof Maziarz *(Jagiellonian University)*

Routing models and how to train them

Standard neural networks apply the same operations to every input example. While this makes them straightforward to train with gradient descent, it doesn't seem optimal - in virtually any domain there is plenty of information to be learned, but a lot of that information is useless for most inputs. Routing models provide a more flexible alternative, where every input is routed through some subgraph of a larger neural network, selectively activating only some of its parameters. This can act as a useful prior and be drastically more efficient, but introduces challenges of its own, when the routing pattern and the network parameters are optimized jointly.

In this talk, we will dive into routing models, and review existing methods backed up by experimental results. I will draw from my research on routing networks performed during my internships at Google in 2016 and 2019, and bridge the three year gap with other related works. Finally, I will give a hint about ongoing efforts and speculate what the future of routing networks might hold.

Maciej Wiatrak *(University of Edinburgh)*

Learning to cooperate with Multi-Agent Reinforcement Learning

Human intelligence is a deeply social phenomenon. Emerging from constant competition and cooperation, it encapsulates years of social interactions. Being innately social animals, our intelligence is centred around interactions and collective behaviour. In order to study the emergence of collective behaviour, this study utilizes multi-agent reinforcement learning to simulate and observe how artificial agents learn to cooperate in order to sustainably manage resources. In other words, the agents powered by novel deep reinforcement learning algorithms are deployed in a game theory interactive dilemma, where they are faced with the challenge of learning how to cooperate, thus preventing the depletion of the resources.

We show how by interacting with the environment through trial-and-error, the agents manage to learn how to successfully manage resources. Thus, proving that advanced human-like cognitive capabilities are not necessarily required to achieve socially desirable outcomes and providing insights to forming rules leading to such equilibria.

The research aims at discovering the underlying reasons behind the interactions between artificial agents. In particular, cooperation and using these insights to contribute to the creation of better mechanism designs, such as international climate protocols. I believe that building and observing the interactions between intelligent systems, such as artificial intelligence agents, could provide us with a way to better understand human behaviour, and in turn, help us construct mechanisms and machines that would assist us in our everyday life, capable of harmonious reasoning and action within the human society.

Maciej Zamorski *(Tooploox / Politechnika Wrocławska)*

Adversarial Autoencoders for Compact Representations of 3D Point Clouds

Deep generative architectures provide a way to model not only images but also complex, 3-dimensional objects, such as point clouds. In this work, we present a novel method to obtain meaningful representations of 3D shapes that can be used for challenging tasks including 3D points generation, reconstruction, compression, and clustering. Contrary to existing methods for 3D point cloud generation that train separate decoupled models for representation learning and generation, our approach is the first end-to-end solution that allows to simultaneously learn a latent space of representation and generate 3D shape out of it. Moreover, our model is capable of learning meaningful compact binary descriptors with adversarial training conducted on a latent space. To achieve this goal, we extend a deep Adversarial Autoencoder model (AAE) to accept 3D input and create 3D output. Thanks to our end-to-end training regime, the resulting method called 3D Adversarial Autoencoder (3dAAE) obtains either binary or continuous latent space that covers a much wider portion of training data distribution. Finally, our quantitative evaluation shows that 3dAAE provides state-of-the-art results for 3D points clustering and 3D object retrieval.

Marcin Tabaka *(Broad Institute of MIT and Harvard)*

Reconstruction of developmental landscapes from single-cell gene expression profiles

Recent progress in developing high-throughput single-cell methods allows researchers to study cell types and states of a tremendous number of cells. Further advances and focused international initiatives, such as the Human Cell Atlas, will likely allow the number of cells that can be analyzed to grow even further, to hundreds of millions of cells and beyond. Deriving biological insights from such massive datasets requires new tools for big data analysis. In particular, understanding the molecular programs that guide differentiation during development is a major challenge. I will describe two projects that tackle different challenges within single-cell data analysis: Waddington-OT and scSVA. My colleagues and I applied Waddington-OT (optimal transport) to reconstruct the landscape of mouse embryonic fibroblasts (MEFs) reprogrammed to induced pluripotent stem cells (iPSCs) from 315,000 single-cell RNA-Seq profiles. The analysis predicts transcription factors and paracrine signals that affect cell fates. scSVA (single-cell Scalable Visualization and Analytics) relies on advances from diverse data-heavy areas, especially astronomy, to scale up most of its capabilities to a billion cells with real time interactivity. To reduce memory usage, scSVA supports efficient retrieval of cell features from massive expression matrices stored on a disk. To facilitate reproducible research, scSVA supports interactive analytics in the cloud with containerized tools. Thus, scSVA should enable users to interact with large datasets and complex analytics to yield novel insights and discoveries.

Marek Rogala *(Appsilon)*

#Ethics - Take responsibility for your model because no one else will

Imagine your disabled grandmother's healthcare plan gets radically cut due to an error in a ML model. She will no longer get the care she needs. This is a life-shattering situation. Unfortunately, this isn't fiction - this has actually happened in 2016 to a woman in Arkansas, US.ML models used for public policy design, law enforcement or insurance have tremendous impact on real human lives. Even when it's not so evident, our models affect humans toa greater extent than we tend to think. For example, a recommendation model at Meetup would be much more likely to recommend tech meetups to men than to women. This would cause a negative feedback loop, further harming diversity. Thanks to a conscious product decision, Meetup team has avoided this bias. In extreme cases, social media recommendation models have already caused unrest and wars in different parts of the world.As Machine Learning engineers, we are best equipped to think of scenarios in which things can go wrong with our model. We're always trying to do the right thing. We also need to be thoughtful enough to foresee and avoid dangers to people affected by our models. We need to dedicate time and energy not only to maximize the metrics, but also to validate our models from ethical perspective. We need to analyze potential risks and convey them to decision makers.This is about human lives and the world we will live in. We must take the responsibility, because no one else will.

Maria Kubara *(University of Warsaw)*

Spatial econometrics with self-made weighting matrixes - uncovering similarity of sample with machine learning results and categorical variables.

Crucial part of spatial econometrics are weighting matrixes. However, spatial dependency is not the only relation, that can be adapted in this form. R package spdep:: provides a method to build own matrixes and convert them to listw class. Therefore, this function opens a possibility to utilize user-build objects in modeling. Filtering not only for geographical dependence, but also for heterogeneity of sample, they can significantly reduce the overbias of standard models. They can be used as an alternative for dummy-variable in OLS and exchange adjacency matrix in Spatial Durbin Model. Using Iris dataset and NUTS4 panel data two case-studies were presented. Categorical variable and machine learning results were used to uncover similarity of data. OLS modeling was augmented with self-made weighting matrixes and, as a result, lowest values of Information Criteria were obtained. The author stressed that weighting matrixes build on categorical data and clustering results can significantly improve econometrical estimation.

Mateusz Buda *(IQVIA / Duke University)*

Radiologist-level deep learning for management of thyroid nodules based on US images

Thyroid nodules are estimated to affect as much as 50% of the population. Triaging them for a biopsy is done based on assessment of ultrasound imaging by radiologists. Our dataset included 2754 US images for 1377 thyroid nodules from 11230 patients. All nodules were assigned ground truth labels (benign or cancer) proven by either biopsy or surgery. In addition, each nodule was assessed for five subjective US imaging feature categories by a single radiologist. For testing, we selected 99 cases that were additionally annotated by multiple radiologists. In a retrospective study, we trained a multi-task deep convolutional neural network to provide biopsy recommendations for thyroid nodules. Auxiliary tasks in the multi-task setting were prediction of the five US imaging feature categories: composition, echogenicity, shape, margin, and calcifications. All tasks were used to train a convolutional network for extraction of a shared feature representation of size 5x5x48. Then, a set of task-specific weights was added and the entire multi-task network was trained end-to-end. Since the distribution of classes in these tasks was imbalanced, we utilized focal loss summed across all tasks with equal weights as an optimization objective. Then, we compared our model to radiologists for discriminating malignant and benign nodules using area under the ROC curve. Within the training set, in 10-fold cross-validation setting, the network achieved an AUC of 0.78 compared with 0.80 for a single expert radiologist. On the 99 test cases, AUC=0.87 for deep learning was comparable to AUC=0.91 for a consensus of three expert radiologists. On the same set, the mean AUC of nine other radiologists was 0.82 with the lowest AUC being 0.76 and the highest 0.85. In summary, a convolutional neural network trained on a limited dataset in multi-task setting achieved performance matching that of radiologists with varying experience for discriminating malignant and benign nodules.

Pawel Skrzypek *(AI Investments)*

Complex AI Forecasting Methods for Investments Portfolio Optimization

Presentation of the first complete AI investment platform. It is based on most innovative AI methods: most advanced neural networks (ResNet/DenseNet, LSTM, GAN autoencoders) and reinforcement learning for risk control and position sizing using Alpha Zero approach.

It shows how the complex AI system which covers both supervised and reinforcement learning could be successfully used to investment portfolio optimization in real time. The architecture of the platform and used algorithms will be presented together with the workflow of machine learning. Also, the real demo of the platform will be shown. The M4 conference winning forecasting method will be presented, with the application to financial time series forecasting.

Paweł Gora *(University of Warsaw)*

Optimizing complex processes using graph neural networks and genetic algorithms

I will present the recent results achieved by the TensorCell team in the research on optimizing complex processes using AI. We focus on optimizing road traffic (solving the so-called Traffic Signal Setting problem) in cities but similar methods may be applied to solving other complex processes, e.g., cancer treatment. We find the optimal settings using metaheuristics such as genetic algorithms and use the simulations (of traffic or cancer growth) to evaluate quality of settings. In order to accelerate evaluations, we train machine learning models as surrogate models able to approximate outcomes of simulations with good accuracy and a few orders of magnitude faster than the simulation.

In the case of optimizing traffic signal settings, we have recently started using as surrogate models graph neural networks built based on a topology of a road network. This approach outperformed all the previous metamodels, especially in terms of the average and the worst case accuracy, e.g., close to the local optima (and this result is essential in the optimization task).

After the models were trained, we tested several approaches to use genetic algorithms, e.g., island genetic algorithms (in which populations evolve mostly separately but exchange the genetic material from time to time) and proximity-based crossover, in which the information about the location of crossroads is taken into account in the optimization procedure.

Piotr Januszewski *(Gdańsk University of Technology)*
Planning in imagination

The aim of this work was to derive from previous work on model learning in complex high-dimensional decision making problems and apply them to planning in complex tasks. Those methods proved to train accurate models, at least in short horizon, and should open a path for application of planning algorithms to problems without access to an accurate simulator i.e. playing Atari 2600 games from pixels, a platform used for evaluation of general competency in artificial intelligence. The goal was to improve data efficiency without loss in performance compared to state-of-the-art model-free methods. This work focused on three benchmarks: an arcade game with dense rewards Boxing, a challenging environment with sparse rewards Freeway and a complex puzzle game Sokoban.

Deriving from state-of-the-art model learning and model-based RL techniques, three architectures are presented: “Original World Models” (OWM), “World Models and AlphaZero” (W+A) and “Discrete PlaNet” (DPN). Despite many difficulties, DPN finally reached a level of performance equal or higher than strong model-free and model-base baselines in low data regime of up to 1M interactions with the real environment of Boxing.

The most challenging part of this work, underestimated at first by the author, was model learning. Current state-of-the-art model learning methods, although report promising results, were not tested for planning with them using search based algorithms, let alone planning and learning.

Neither architecture was able to learn playing Sokoban. Certainly, the problem lies in model learning techniques. Sokoban dynamics, although based on simple rules of moving a character and pushing boxes, allow for incredible number of possible states and levels configurations. The models were not able to generalize well to this number of possibilities.

In the talk/poster the author will present these and other insight into model-based RL field. This work is author's master thesis topic defended with honours.

Piotr Kozakowski *(Google Brain / University of Warsaw)*
Using Transformers to teach Transformers how to train Transformers

Well-performing deep learning models have enormous impact, but getting them to perform well is complicated - a number of hyperparameters need to be tuned, which requires time-consuming experimentation. Deep learning practitioners commonly keep most of the hyperparameters constant throughout training (momentum, dropout rate), while varying some of them according to a static schedule (learning rate). Some use inventive heuristics to control the hyperparameters dynamically, such as “drop the learning rate two times when the validation accuracy stops increasing”. Inventing such rules is tedious and fails to account for more complex interactions between hyperparameters.

We propose a reinforcement learning-based method of learning hyperparameter controllers which, based on the current values of relevant metrics, tune the hyperparameters to optimize the final model performance. Since data for controller training is very expensive, we introduce a Transformer-based forecaster network, predicting a sequence of training metrics based on the actions performed by the controller. We train the controller using data generated by the forecaster network instead of real runs of model training, achieving a massive speed-up of 1000x in GPU-time, while still being able to learn effective controller policies.

I will start the talk by describing the problem of adaptive hyperparameter tuning and its solution using vanilla reinforcement learning. I will proceed to introducing the training routine for the forecaster network and showcasing its ability to model time series on both synthetic and real data. Then I will present the results of the experiments, comparing the performance of the learned controllers against human-tuned baselines on various tasks. I will end with a qualitative analysis of the controllers and sharing insights from the hyperparameter schedules generated by them.

Work has been done during an internship at Google Brain and is currently under review for ICLR 2020.

Piotr Zioło *(ORA AI)*

Automated decision making in online advertising

To effectively manage Google Ads campaigns one has to make hundreds of decisions a day. The advertiser has to set the amounts he is willing to pay for a click for every of the campaign's keywords, as well as for each of the available segments (devices, gender, age etc.). Doing that he has to take into account the characteristics of his business, time-related factors and his competitors. The environment a campaign runs in evolves quickly as other advertisers either use skills of digital marketing managers or advanced algorithms. Therefore, the only viable option to effectively use the budget for online advertising is to switch to automated bidding algorithms.

There are several families of decision-making algorithms. A huge dataset which does not outdate quickly can be used to train DNNs to do the job. This might be a viable option for Google, but not so much for other advertisers who only have their own data. Moreover, even for Google it is not optimal because of the fast-changing reality. Another option is to use active learning algorithms, which aim to optimize profit and learn at the same time. One family of such algorithms stems from computer science and its most prominent example is UCB. The other family of algorithms involves Bayesian optimization and the go-to algorithms are the Expected Improvement and Knowledge Gradient.

In this talk I will describe a methodology how to properly test the effectiveness of decision-making algorithms and the stochastic simulator we developed to test bidding algorithms for Google Ads. This simulator allows to realistically emulate Google Ads, including non-stationarity and adversarial behaviors. Then I will give a brief overview of known active learning algorithms and show their performance in the simulator. Because most algorithms described in literature assume stationarity, they do not fare well in reality. Finally, I will describe the approaches we took to actually make those algorithms work well in practice.

Przemysław Ryś *(University of Warsaw)*

Markov Chain Monte Carlo methods in Machine Learning

Markov Chain Monte Carlo (MCMC) is a class of statistical algorithms for sampling from a complicated probability distribution in an approximate way. The approach was revolutionary for applications of Bayesian inference, allowing to construct more complex, realistic models. They become also foundations of widely used machine learning algorithms. My presentation (or poster) would include an explanation of construction and idea of MCMC algorithms such as Metropolis-Hastings and Gibbs Sampler with special focus on their mathematical foundations. I would like to also present briefly examples from a wide landscape of machine learning methods based on MCMC, such as latent Dirichlet allocation, hierarchical classification or applications in deep learning and image modeling. Markov chain based algorithms are also helpful in global optimizations problems - even stochastic gradient method used to fit non-linear models is a special case of Markov Chain Monte Carlo. One of the examples is the algorithm, designed by me, which fits derivatives pricing model, based on latent stochastic process, to market data using Bayesian inference.

Stanisław Jastrzębski *(New York University)*

Gradient alignment: understanding when stochastic gradient descent works, and when it doesn't

Deep neural networks are almost always trained using a simple gradient-based method such as Adam or SGD. Interestingly, in deep learning optimization and generalization cannot be analysed separately. For instance, using a large learning rate is often critical for achieving good generalization. In the first part of the talk I will argue that a simple concept called gradient alignment helps explain multiple puzzles about optimization in deep learning such as the role of learning rate in generalization or how deep networks avoid "memorizing" examples. In the the second part, I will show that gradient alignment provides a new perspective on currently open challenges in deep learning such as multi-task learning or multi-modal learning.

Tomasz Konopczyński

(Tooploox / Heidelberg University)

Plugin Networks for Inference under Partial Evidence

In this work we propose a novel method to incorporate partial evidence in the inference of deep convolutional neural networks. Contrary to the existing, top performing methods, which either iteratively modify the input of the network or exploit external label taxonomy to take the partial evidence into account, we add separate network modules (”Plugin Networks”) to the intermediate layers of a pretrained convolutional network. The goal of these modules is to incorporate additional signal, i.e. information about known labels, into the inference procedure and adjust the predicted output accordingly. Since the attached plugins have a simple structure, consisting of only fully connected layers, we drastically reduced the computational cost of training and inference. At the same time, the proposed architecture allows to propagate information about known labels directly to the intermediate layers to improve the final representation. Extensive evaluation of the proposed method confirms that our Plugin Networks outperform the state-of-the-art in a variety of tasks, including scene categorization, multi-label image annotation and semantic segmentation.

Wojciech Rządkowski *(Institute of Science and Technology Austria (IST Austria))*

Machine learning & quantum physics: how convolutional neural networks discover phases of matter

Recently, machine learning tools are quickly gaining popularity among natural scientists. Conveniently programmable using modern libraries, they can learn phases of matter, automatically design quantum experiments, or assist in adaptive measurement, to name just a few applications in physics and chemistry.

We shall briefly discuss various applications of machine learning in physics. Afterwards, we introduce our recent work on discovering phases of matter in multi-layered models using convolutional neural networks. In particular, we demonstrate how our unsupervised algorithm can learn inter-layer correlations, "hidden" to visual inspection of the system samples, are learned.

The contribution will be largely based on our recent preprint "Detecting hidden and composite orders in layered models via machine learning", https://arxiv.org/abs/1907.05417. Strong emphasis will be put, however, on making the content accessible to non-physicists.