Theoretical Foundations of
Machine Learning
11-15 February 2019
Jagiellonian University, Faculty of Mathematics and Computer Science
Cracow, Łojasiewicza 6

LECTURES
Andreas Bender, University of Cambridge
Wojciech Czarnecki, DeepMind
Asja Fischer, Ruhr-University Bochum
Krzysztof Geras, New York University
Emtiyaz Khan, RIKEN
Razvan Pascanu, DeepMind
Ilya Tolstikhin, Google AI

KEY DATES
research paper submission - 30 Nov 2018
extended abstract submission - 21 Dec 2018
authors notification - 05 Jan 2019
registration 5-19 Jan 2019

Organizers
Faculty of Mathematics
and Computer Science
of the Jagiellonian University

www: tfml.gmum.net
e-mail: tfml@uj.edu.pl
Date

TFML 2019 conference will take place at the Faculty of Mathematics and Computer Science, Jagiellonian University, Kraków, Poland, on February 11-15, 2019.

Main goals

The aim of the conference is to integrate the machine learning community, popularize new algorithms and methods in the field, and provide the opportunity to exchange the experience among the participants.

Scope

The conference scope covers wide area of machine learning topics, including, but not limited to active learning, representation learning, semi-supervised learning, supervised learning, unsupervised learning, cheminformatics, computational complexity, deep learning, adversarial networks, attention models, computer vision, generative models, natural language processing, speech recognition, cognitive science, optimization methods, and reinforcement learning. Both, from theoretical and practical point of view.

Organizers

Sponsors

Gold sponsor

Silver sponsors
Place

Address:
ul. Łojasiewicza 6
30-348 Krakow

Location: All presentations will take place in the lecture hall 0089 (marked in blue on the building plan below).
<table>
<thead>
<tr>
<th>Time</th>
<th>Monday</th>
<th>Tuesday</th>
<th>Wednesday</th>
<th>Thursday</th>
<th>Friday</th>
</tr>
</thead>
<tbody>
<tr>
<td>09:15</td>
<td>Plenary Lecture 2 Krzysztof Geras</td>
<td>Plenary Lecture 4 Razvan Pascanu</td>
<td>Plenary Lecture 6 Thomas Kipf</td>
<td>Plenary Lecture 8 Emtiyaz Khan</td>
<td></td>
</tr>
<tr>
<td>10:30</td>
<td>Coffee Break</td>
<td>Coffee Break</td>
<td>Coffee Break</td>
<td>Coffee Break</td>
<td>Coffee Break</td>
</tr>
<tr>
<td>11:00</td>
<td>Registration</td>
<td>Session 2</td>
<td>Sponsors Session</td>
<td>Session 4</td>
<td>Session 6</td>
</tr>
<tr>
<td>13:00</td>
<td>Lunch</td>
<td>Lunch</td>
<td>Lunch</td>
<td>Lunch</td>
<td>Lunch</td>
</tr>
<tr>
<td>14:00</td>
<td>Opening Ceremony</td>
<td>Plenary Lecture 1 Wojciech Czarnecki</td>
<td>Plenary Lecture 3 Illya Tolstikhin</td>
<td>Plenary Lecture 5 Asja Fischer</td>
<td>Session 5</td>
</tr>
<tr>
<td>15:15</td>
<td>Coffee Break</td>
<td>Coffee Break</td>
<td>Poster Session / Coffee Break</td>
<td>Coffee Break</td>
<td>Coffee Break</td>
</tr>
<tr>
<td>16:00</td>
<td>Session 1</td>
<td>Event</td>
<td>Session 3</td>
<td>Plenary Lecture 7 Andreas Bender</td>
<td></td>
</tr>
<tr>
<td>18:00</td>
<td>Gala Dinner</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>19:00</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
PLENARY LECTURES
Chair: Jacek Tabor, Igor Podolak

1 Wojciech Czarnecki (DeepMind)
   Beyond First-order Optimisation - Population-based and Open-ended Learning

2 Krzysztof Geras (New York University)
   Solving breast cancer screening diagnosis with deep learning

3 Ilya Tolstikhin (Google AI)
   Wasserstein Auto-Encoders: Basics, Design Choices, and Challenges

4 Razvan Pascanu (DeepMind)
   Looking at transfer in reinforcement learning settings

5 Asja Fischer (Ruhr-University Bochum)
   Various aspects of generative models

6 Thomas Kipf (University of Amsterdam)
   Learning with graph-structured representations

7 Andreas Bender (University of Cambridge)
   Machine Learning in Drug Discovery - From Hype to Reality

8 Emtiyaz Khan (RIKEN)
   Fast yet Simple Natural-Gradient Descent for Variational Inference in Deep Learning

LEGEND
- 60 minutes talk and 10 minutes for questions and comments
- 25 minutes talk and 5 minutes for questions and comments
- 15 minutes talk and 5 minutes for questions and comments
- 10 minutes talk and 5 minutes for questions and comments

SESSION 1
Chair: Jacek Tabor

1 Pawel Gora (University of Warsaw)
   Investigating performance of neural networks and gradient boosting models approximating microscopic traffic simulations in traffic optimization tasks

2 Dániel Varga, Adrián Csiszárik and Zsolt Zombori
   Gradient Regularization Improves Accuracy of Discriminative Models

3 Tomasz Kisielewski and Damian Leśniak
   Accidental exploration through value predictors

4 Antreas Antoniou, Agnieszka Słowiak, Elliot Crowley, Amos Storkey
   Dilated DenseNets for Relational Reasoning

5 Mateusz Staniak, Przemysław Biecek
   In Search of Interpretable Features to Explain Decisions of Black-Box Models

6 Katarzyna Janocha
   Learning to be Interpretable
### SESSION 2  
**Chair:** Marek Śmieja

| 1 | Mikofaj Bińkowski (Imperial College London)  
*Better GANs with MMD* |
| 2 | Maciej Wołczyk  
*Deep learning-based initialization for object packing* |
| 3 | Bartosz Wojcik, Łukasz Maziarzka, Jacek Tabor  
*LOSSGRAD: automatic learning rate in gradient descent* |
| 4 | Szymon Knop, Marcin Mazur, Jacek Tabor, Igor Podolak, Przemysław Spurek  
*Comparison of auto-encoder sliced generative models* |
| 5 | Beatrix Benkő, Adrien Csizárak  
*One-class Classification with a Hybrid VAE-GAN Model* |
| 6 | Agnieszka Pocha, Krzysztof Hajto, Krzysztof Misztal  
*Does FWL-teacher has to be a Gaussian Process?* |

### SPONSOR SESSION  
**Chair:** TBA

| 1 | Oleksandr Zakharchuk  
*Detecting duplicate content in Brainly* |
| 2 | Jan Kaczmarczyk  
*Methods for high-dimensional data: example of DREAM Proteogenomics* |
| 3 | Jan Malolepsy  
*About Pegasystems – the leader in BPM platform* |

### POSTER SESSION

| 1 | Paulina Kozioł, Magda Raczkowska, Danuta Libera, Slawka Urbaniak, Czesława Paluszkiewicz, Wojciech Kwiatek, Tomasz Wrobel.  
*Random forest classifier potential in the pancreatic cancer diagnosis with FT-IR imaging* |
| 2 | Agnieszka Kraft, Ziad Al Bkhetan, Michał Kadłoś, Dariusz Plewczynski  
*Inferring Three-Dimensional Chromatin Organization of Human Cells from Machine Learning Predicted Chromatin Interactions* |
| 3 | Piotr Woźniak, Michał Kusza, Piotr Migdał  
*Modelling response to trypophobia trigger using intermediate layers of ImageNet networks* |
| 4 | Łukasz Mądry  
*Detecting out of distribution samples with random network distillation* |
| 5 | Agnieszka Pocha, Krzysztof Hajto, Krzysztof Misztal  
*Does FWL-teacher has to be a Gaussian Process?* |
| 6 | Marcin Kowalik  
*Estimation of Critical Temperature of high-temperature Superconductor by Neural Network* |
| 7 | Krzysztof Rajda, Przemysław Spurek, Jacek Tabor  
*GMM CWAE* |
| 8 | Tomasz Danel  
*Lichen Classifier Leveraging Recent Advances in Mobile Neural Networks* |
| 9 | Przemysław Spurek, Łukasz Maziarzka, Aleksandra Nowak, Jacek Tabor, Stanisław Jastrzębski  
*Non-linear ICA based on Cramer-Wold metric* |
| 10 | Stanisław Jastrzębski, Maciej Szymczak, Agnieszka Pocha, Sabina Podlewska  
*How to dock without docking - prediction of interaction fingerprints from 2D compound structure* |
| 11 | Michał Makowski  
*Precision matrix estimation in sparse Gaussian graphical models gLasso and gSLOPE approach* |
SESSION 3
Chair: Krzysztof Misztal
1. Piotr Warchol (Jagiellonian University)
   Dynamical Isometry is Achieved in Residual Networks in a Universal Way for any Activation Function
2. Michał Łukasik (Google AI)
   Content Explorer: Recommending Novel Entities for a Document Writer

SESSION 4
Chair: Igor Podolak
1. Sabina Podlewska (Institute of Pharmacology of the Polish Academy of Sciences / Jagiellonian University Medical College)
   Machine learning in the serve of searching for new drugs
2. Stanisław Jastrzębski, Maciej Szymczak, Agnieszka Pocha, Sabina Podlewska
   How to dock without docking - prediction of interaction fingerprints from 2D compound structure
3. Monika Piwowar, Jacek Dygut, Piotr Piwowar
   The GCA/ GCCA methods - a way of extracting genetic signatures from gene expression arrays
4. Magdalena Wiercioch
   3DST, A Novel Molecule Representation
5. Danuta Liberda, Michael Hermes, Paulina Koziol, Magda Raczkowska, Nicholas Stone, Tomasz Wróbel
   Impact of paraffin presence on Random Forest classification of FTIR hyperspectral images of esophagus cancer
6. Mateusz Półtorak, Mikołaj Morzy
   Emotion Recognition in Raw Speech Recordings Using Convolutional Neural Networks and Spectrograms

SESSION 5
Chair: Bartosz Zieliński
1. Tomasz Trzciński (Tooploox / Warsaw University of Technology), Maciej Zięba (Tooploox / Warsaw University of Technology)
   BinGAN: Learning Compact Binary Descriptors with a Regularized GAN
3. Michał Zając, Konrad Żołna, Negar Rostamzadeh, Pedro Pinheiro – Adversarial Framing for Image and Video Classification
4. Marcin Możęjko, Michał Zmysłowski
   Penalizing approximated Hessian norm against adversarial attacks
5. Michał Sadowski, Aleksandra Grzegorczyk
   Image Inpainting with Gradient Attention
6. Krzysztof Misztal, Przemysław Spurek, Jacek Tabor
   Image Stitching Based on Entropy Minimization

SESSION 6
Chair: Jacek Tabor
1. Jarek Duda (Jagiellonian University)
   Hierarchical correlation reconstruction - between statistics and ML
2. Marcin Mazur, Piotr Kościelnia
   On some goodness of fit tests for normality based on the optimal transport distance
3. Jan Kocoń
   Improved recognition and normalisation of Polish temporal expressions using Cascade of Partial Rules
4. Jakub Chłędowski, Tomasz Wesołowski, Stanisław Jastrzębski
   Improving Utilization of Lexical Knowledge in Natural Language Inference
Igor Sieradzki, Igor Podolak

*Distribution-Interpolation Trade off in Generative Models*
TFML 2019: Plenary lectures
Wojciech Czarnecki  
DeepMind

Wojciech obtained his PhD at the Jagiellonian University, under supervision of prof. Jacek Tabor for thesis at the intersection of Machine Learning and Information Theory. Since then he has been working at DeepMind, on problem spanning from theoretical machine learning, through optimisation techniques, to deep reinforcement learning and most recently multi-agent deep reinforcement learning.

Plenary lecture 1: *Beyond First-order Optimisation - Population-based and Open-ended Learning*

**Abstract:** Gradient based optimisation became a core tool of modern machine learning systems. While differentiable loss functions can describe wide range of problems and phenomena from image recognition through audio synthesis to medical predictions systems, there are many challenges in artificial intelligence that lack these properties. Various competitive and cooperative behaviours such as the ones emerging when playing multiplayer games, or in general-in multi agent systems, cannot be represented as a simple optimisation problem. In this talk we will take a closer look at going beyond single function minimisation - towards training populations of agents with both well specified goals as well as in completely open-ended fashion, where there is no objective notion of "better" agents/models, but only pair-wise, highly non-transitive, competition.

*Relevant papers:*
1. Open-ended Learning in Symmetric Zero-sum Games
2. Evolving intrinsic motivations for altruistic behavior
3. Human-level performance in first-person multiplayer games with population-based deep reinforcement learning

Krzysztof Geras  
New York University

Krzysztof is an assistant professor at NYU. His main interests are in unsupervised learning with neural networks, model compression, transfer learning, evaluation of machine learning models and applications of these techniques to medical imaging. He previously did a postdoc at NYU with Kyunghyun Cho, a PhD at the University of Edinburgh with Charles Sutton and an MSc as a visiting student at the University of Edinburgh with Amos Storkey. His BSc is from the University of Warsaw. He also did industrial internships in Microsoft Research (Redmond, working with Rich Caruana and Abdel-rahman Mohamed), Amazon (Berlin, Ralf Herbrich's group), Microsoft (Bellevue) and J.P. Morgan (London).

Plenary lecture 2: *Solving breast cancer screening diagnosis with deep learning*

**Abstract:** Although deep learning has made a stunning progress in the last few years, both in terms of engineering and theory, its real-life applications remain rather limited. One of the fields that has been anticipated to be revolutionized by deep learning for some time, yet proved to be much harder that many expected, is medical diagnosis. In this talk I will shed some light on my 2.5-year long journey in developing deep learning methods for medical imaging, in particular, for breast cancer screening. I will explain how we created a deep learning model that can perform a diagnosis with an accuracy comparable to experienced radiologists. To achieve this goal we needed a lot of perseverance, novel neural network architectures and training methods specific to medical imaging. I will also discuss the limitations of our work and what can be likely achieved in the next few years.
Ilya Tolstikhin
Google AI

Ilya Tolstikhin completed his PhD in 2014 in Moscow, Russia, working with Konstantin Vorontsov on statistical learning theory. He spent 4 years as a postdoc in Bernhard Schoelkopf’s lab at Max Planck Institute for Intelligent Systems, Tuebingen, Germany, where he continued his research on learning theory and also developed an interest in unsupervised generative modeling. In 2018 he joined Brain Team at Google AI, Zurich as a research scientist.

Plenary lecture 3: Wasserstein Auto-Encoders: Basics, Design Choices, and Challenges

Abstract: This talk will focus on the Wasserstein auto-encoders (WAE) which is a recently introduced approach to the unsupervised generative modeling. WAE shares many of the nice properties of VAEs (stable training, encoder-decoder architecture) while generating samples of better quality, as measured by the FID scores. The talk will also highlight some of the more recent developments related to WAEs including various choices of the regularizer and curious effects of the latent space dimensionality on the latent representations.

Razvan Pascanu
DeepMind

Razvan Pascanu is a research scientist at DeepMind in London. He obtained his bachelor and master degree in Germany, at Jacobs University, and the Ph.D. at Université de Montreal, working with Yoshua Bengio on optimization, recurrent models and deep learning in general. His interests range from topics like optimization, neural networks to deep reinforcement learning, continual learning and structured neural networks models.

Plenary lecture 4: Looking at transfer in reinforcement learning settings

Abstract: Deep Reinforcement Learning, while providing some impressive results (e.g. on Atari, Go, etc.), is notoriously data inefficient. This is partially due to the function approximators used (deep networks) but also due to the weak learning signal (based on observing rewards). In this talk we will discuss the role of transfer learning to help making DRL more data efficient. In particular I will focus on how different formulation of KL-regularized RL can provide more systematic exploration of the environment and hence a more reliable learning signal. If time allows, we will quickly cover three related recent works https://arxiv.org/abs/1707.04175, https://arxiv.org/abs/1806.01780 and recent work on information asymmetry in KL-regularized settings.
Asja Fischer  
Ruhr-University Bochum

Asja is an assistant professor heading the Machine Learning group at the Faculty of Mathematics at the Ruhr University in Bochum. From 2016 to 2018 she was assistant professor at the University of Bonn and a post-doctoral researcher at the Montreal Institute for Machine Learning in 2015. Between 2010 and 2015, she was employed both at the Institute for Neural Computation, Ruhr-University Bochum, Germany, and the Department of Computer Science, University of Copenhagen, Denmark, working on her PhD, which she defended in Copenhagen in 2014. Before, she studied Biology. Bioinformatics, Mathematics and Cognitive Science at the Ruhr-University Bochum, the Universidade de Lisboa and the University of Osnabrück.

Plenary lecture 5: Various aspects of generative models

Abstract: While supervised deep learning lead to breakthroughs in various areas of applications, the development and understanding of unsupervised generative models constitute an active area of research. In this talk I will discuss two kind of generative models: 1) Restricted Boltzmann machines, which are undirected generative models that require Markov chain Monte Carlo sampling for approximating the log-likelihood gradient during training. 2) Wasserstein generative adversarial networks (WGANs), that belong to the class of directed generative models and apply adversarial learning to minimize the Wasserstein distance, which introduces a Lipschitz constraint into the optimization problem. Specifically, I will discuss different sampling strategies for RBMs, and a way to enforce the Lipschitz constraint in WGANs.

Thomas Kipf  
University of Amsterdam

Thomas Kipf is a third-year PhD student at the University of Amsterdam advised by Max Welling. His research focuses on deep learning with (graph-)structured representations, including topics such as semi-supervised learning, multi-agent modeling, and structured deep generative models. He has made several contributions in the field of graph representation learning and is the lead developer of Graph Convolutional Networks and Graph Auto-Encoders, popular frameworks for both supervised and unsupervised learning with graph-structured data.

Plenary lecture 6: Learning with graph-structured representations

Abstract: In recent years there has been an increasing number of success stories in applying ideas from deep learning to graph-structured data. The workhorse in this emerging field is the graph neural network: a message passing algorithm parameterized by neural networks, trained via backpropagation. Variants of graph neural networks now define the state of the art in many classical graph or network problems, such as node classification, graph classification, and link prediction. In this talk, I will give an overview of structured deep models that employ graph neural networks as a key component and discuss trade-offs for a few popular model variants such as graph convolutional networks (GCNs) [1] and message passing neural networks [2]. I will further introduce two emerging research directions: learning deep generative models of graphs and inference of latent graph structure. Structured deep models are ideal candidates for these areas and hold great promise for applications such as chemical synthesis, relation extraction in NLP, and in modeling the physical world.

Andreas Bender  
University of Cambridge  

Dr Andreas Bender is a Reader for Molecular Informatics with the Centre for Molecular Science Informatics at the Department of Chemistry of the University of Cambridge, leading a group of about 22 postdocs, PhD and graduate students and academic visitors. In his work, Andreas is involved with the integration and analysis of chemical and biological data, aimed at understanding phenotypic compound action (such as cellular readouts, and also organism-level effects) on a mechanistic level, predicting molecular properties related to both compound efficacy and toxicity, as well as drug repurposing. He received his PhD from the University of Cambridge and worked in the Lead Discovery Informatics group at Novartis in Cambridge/MA as well as at Leiden University in the Netherlands before his current post. In 2013 he was awarded an ERC Starting Grant to model mixture effects of chemical structures in biological systems using mechanistic approaches, an area currently very little understood.

Plenary lecture 7: Machine Learning in Drug Discovery - From Hype to Reality

Abstract: In this presentation, which is split into two parts, we will firstly (by Andreas Bender) outline which types of data can be used in the drug discovery context to support both the discovery of New Medical Entities (NMEs), and the repurposing of existing drugs. Examples from analyzing the mode of action of compounds to the selection of selective or ‘selectively promiscuous’ compounds will be given, as well as the utilization of gene expression data and knowledge graphs for drug repurposing. This part of the contribution will conclude with a ‘reality check’ of the opportunities (and pitfalls) of using machine learning and ‘AI’ in the drug discovery context. In the second part (by Isidro Cortes Ciriano) novel applications of machine learning approaches in the drug discovery field will be presented, that encompass both aspects related to the representation of data, as well as learning in more efficient ways from chemical and biological information sources. In particular, the representation of molecular structures in 'graphical' and fingerprint formats will be explored, as well as ways of estimating errors of neural networks in bioactivity models from ensemble snapshots, among other topics.

Emtiyaz Khan  
RIKEN  

I am a team leader at the RIKEN center for Advanced Intelligence Project (AIP) in Tokyo where I lead the Approximate Bayesian Inference (ABI) Team. From April 2018, I am a visiting professor at the EE department in Tokyo University of Agriculture and Technology (TUAT). I am an Action Editor for the Journal of Machine Learning (JMLR). From 2014 to 2016, I was a scientist at EPFL in Matthias Grossglausser’s lab. During my time at EPFL, I taught two large machine learning courses for which I received a teaching award. I first joined EPFL as a post-doc with Matthias Seeger in 2013 and before that I finished my PhD at UBC in 2012 under the supervision of Kevin Murphy.

Plenary lecture 8: Fast yet Simple Natural-Gradient Descent for Variational Inference in Deep Learning

Abstract: Variational Inference plays an important role in advancing machine learning and deep learning methods, but it typically requires more effort to implement and execute compared to maximum-likelihood methods. In this talk, I will summarize some of our recent work that enables fast and yet simple inference for deep models, such as Bayesian neural network. I will show that natural-gradients play a fundamental role combining many existing methods in optimization and statistics (e.g., EM algorithm and message passing algorithms), and surprisingly they are also simple to implement for some problems. I will end with a summary of some of the current challenges that we face in variational inference for deep learning.
TFML 2019: Sponsor session
1  Oleksandr Zakharchuk  
**Brainly**

**Bio:** For the last two years Sashko been advising startups on applying Deep Learning, mostly in the domains of Natural Language Processing and Recommender Systems. Previously, he worked at Google, building models for personalized content recommendations, systems for exploring new content, as well as compact language models.

**Title:** *Detecting duplicate content in Brainly*

**Abstract:** In this presentation, we’ll cover some practical aspects of applying ML in startup environment, and illustrate them with examples from duplicate content finder project in Brainly.

---

2  Jan Kaczmarszyczyk  
**Ardigen**

**Bio:** Jan Kaczmarszyczyk completed his PhD in Theoretical Physics in 2011 at the Jagiellonian University. After doing two postdocs, he switched field to Machine Learning in 2017. Since then he has been working at Ardigen S.A., where he specializes in Machine Learning for Life Sciences, including Genomics, Immuno-Oncology, and Computer Aided Drug Design

**Title:** *Methods for high-dimensional data: example of DREAM Proteogenomics Challenge*

**Abstract:** TBA

---

3  Jan Małolepszy  
**Pegasystems**

**Bio:** MS in Computer Science from Jagiellonian University in 1996. Worked for a number of Polish and international companies delivering business solutions for corporate market. Since 2013 leads the Engineering team in Pegasystems office in Kraków.

**Title:** *About Pegasystems – the leader in BPM platform*

**Abstract:** Pegasystems, the leader in software for customer engagement and operational excellence. Pega Infinity - adaptive, cloud-architected software – built on the unified Pega® Platform – empowers people to rapidly deploy, and easily extend and change applications to meet strategic business needs.
TFML 2019: Sessions
SESSION 1

1. Paweł Góra
   University of Warsaw

   Investigating performance of neural networks and gradient boosting models approximating microscopic traffic simulations in traffic optimization tasks

   Abstract: Vehicular traffic in cities is a very complex phenomenon and it is not easy to manage it properly. In the TensorCell project, we train metamodels, such as neural networks and gradient boosting models, to approximate outcomes of microscopic traffic simulations very fast and with good accuracy. Thanks to that, we can apply evolutionary algorithms to find suboptimal settings of different traffic management strategies, e.g., traffic signal control settings - such configurations can be efficiently evaluated using the trained metamodels. However, it turns out that the accuracy of metamodels may decrease close to local optima which makes the optimization task more difficult. I will talk about results of analysis of this phenomenon, in which many metamodels and settings of genetic algorithms were investigated in order to understand what are the reasons of this phenomenon, what is its scale, how it can be mitigated and what can be potentially done to design better real-time traffic optimization methods.

2. Piotr Warchoł
   Jagiellonian University

   Dynamical Isometry is Achieved in Residual Networks in a Universal Way for any Activation Function

   Abstract: We demonstrate that in residual neural networks (ResNets) dynamical isometry is achievable irrespectively of the activation function used. We do that by deriving, with the help of Free Probability and Random Matrix Theories, a universal formula for the spectral density of the input-output Jacobian at initialization, in the large network width and depth limit. The resulting singular value spectrum depends on a single parameter, which we calculate for a variety of popular activation functions, by analyzing the signal propagation in the artificial neural network. We corroborate our results with numerical simulations of both random matrices and ResNets applied to the CIFAR-10 classification problem. Moreover, we study the consequence of this universal behavior for the initial and late phases of the learning processes. We conclude by drawing attention to the simple fact, that initialization acts as a confounding factor between the choice of activation function and the rate of learning.

3. Dániel Varga, Adrián Csiszárik and Zsolt Zombori

   Gradient Regularization Improves Accuracy of Discriminative Models

   Abstract: Regularizing the gradient norm of the output of a neural network is a powerful technique, rediscovered several times. This paper presents evidence that gradient regularization can consistently improve classification accuracy on vision tasks, using modern deep neural networks, especially when the amount of training data is small. We introduce our regularizers as members of a broader class of Jacobian-based regularizers. We demonstrate empirically on real and synthetic data that the learning process leads to gradients controlled beyond the training points, and results in solutions that generalize well.
**Tomasz Kisielewski and Damian Leśniak**

**Accidental exploration through value predictors**

**Abstract:** Infinite length of trajectories is an almost universal assumption in the theoretical foundations of reinforcement learning. In practice learning occurs on finite trajectories. In this paper we examine a specific result of this disparity, namely a strong bias of the time-bounded Every-visit Monte Carlo value estimator. This manifests as a vastly different learning dynamic for algorithms that use value predictors, including encouraging or discouraging exploration. We investigate these claims theoretically for a one dimensional random walk, and empirically on a number of simple environments. We use GAE as an algorithm involving a value predictor and evolution strategies as a reference point.

**Antreas Antoniou, Agnieszka Słowińska, Elliot Crowley, Amos Storkey**

**Dilated DenseNets for Relational Reasoning**

**Abstract:** Despite their impressive performance in many tasks, deep neural networks often struggle at relational reasoning. This has recently been remedied with the introduction of a plug-in relational module that considers relations between all pairs of objects. Unfortunately, this is combinatorially expensive. In this work it was shown that a DenseNet model incorporating dilated convolutions excels at relational reasoning on the Sort-of-CLEVR dataset, which allows to forgo the relational module and its associated expense.

**Mateusz Staniak, Przemysław Biecek**

**In Search of Interpretable Features to Explain Decisions of Black-Box Models**

**Abstract:** The LIME algorithm is a popular method of explaining individual decisions of Machine Learning models. LIME fits a simple model locally to the predictions of a complex model. It was successfully applied to models for text and image data. This approach identifies relevant interpretable features (for example parts of an image) that explain the model behavior. For tabular data, the concept of an interpretable feature is not as clear and simple methods such as building new features based on quantiles of the distribution of a variable seem insufficient. In this paper, we propose a new method that explains the behavior of a model built on tabular data for a single observation.

**Katarzyna Janocha**

**Learning to be Interpretable**

**Abstract:** Analysing models is hard. However, it can be treated as a multitask problem - solving original task, learning to interpret and to be interpretable. We propose a novel method of training latent attention networks jointly with the analysed model, obtaining more detailed attention masks than under the regime in which both models are trained separately. The method has regularization properties, resulting in improved generalization, suggesting the model focuses on key features.
1  **Mikołaj Bińkowski**  
Imperial College London  

*Better GANs with MMD*

**Abstract:** Maximum Mean Discrepancies (MMD) form a well-studied subclass of Integral Probability Metrics that quantify differences between probability distributions. Coupled with adversarial training, they provide a powerful objective function for generative modelling. In this talk we focus on recent developments in MMD GANs, including theoretical relations to other GAN objectives, the issue of training bias and regularization. We show that controlling the gradient of the critic is vital to having a sensible loss function. Finally, we will show how to construct a new GAN objective (Scaled MMD) that enforces exact, analytical gradient constraints, and does so at no additional cost as compared to existing approximate techniques based on additive regularizers.

2  **Maciej Wołczyk**  

*Deep learning-based initialization for object packing*

**Abstract:** One of the most important optimization tasks in the industry at the current time is the object packing problem. Although several methods have been developed for the purpose of solving it, they are usually only able to optimize placement locally and as such are heavily dependent on the choice of the initial setting – hence the need for trying out multiple possible starting points, which impacts algorithm running time. In this paper we present a neural network-based model which provides sensible starting points in a linear time.

3  **Bartosz Wojcik, Łukasz Maziarka, Jacek Tabor**  

*LOSSGRAD: automatic learning rate in gradient descent*

**Abstract:** In this paper we propose a simple, fast and easy to implement algorithm LOSSGRAD (locally optimal step-size in gradient descent), which automatically modifies the step-size in gradient descent during learning. Given a function $f$, a point $x$, and the gradient $\nabla f$ of $f$, we aim to find the step-size $h$ which is (locally) optimal, i.e. satisfies:

$$h = \arg\min_{t \geq 0} f(x - t\nabla f).$$

Making use of quadratic approximation, we show that the algorithm satisfies the above assumption. We experimentally show that our method is insensitive to the choice of initial learning rate while achieving results comparable to other methods.
4 Szymon Knop, Marcin Mazur, Jacek Tabor, Igor Podolak, Przemysław Spurek

Comparison of auto-encoder sliced generative models

Abstract: In this paper we discuss a class of AutoEncoder based generative models based on one dimensional sliced approach. The idea is based on the reduction of the discrimination between samples to one-dimensional case. Our experiments show that methods can be divided into two groups. First consists of methods which are a modification of standard normality tests, while the second is based on classical distances between samples. It turns out that both groups are correct generative models, but the second one gives a slightly faster decrease rate of Fréchet Inception Distance (FID).

5 Beatrix Benkő, Adrián Csiszárik

One-class Classification with a Hybrid VAE-GAN Model

Abstract: We propose a hybrid of Variational Autoencoder (VAE) and Generative Adversarial Network (GAN) models for the one-class classification task. The clear probabilistic framework of the VAE model allows studying the problem with an information-theoretic toolset while the powerful generative capabilities of the GAN model enrich the feature set used for the one-class discrimination. Our preliminary results already show improvements over the state-of-the-art results on several classes of the CIFAR-10 dataset.

6 Agnieszka Pocha, Krzysztof Hajto, Krzysztof Misztal

Does FWL-teacher has to be a Gaussian Process?

Abstract: Labels for biological datasets are expensive to obtain. One solution is to use models that do not require training (ex. active contours are used for finding objects’ outline). However, these methods are unable to profit from available data and might require manual tuning for each new task. These obstacles can be overcome with deep learning models, though the question arises: How should the information from the little amount of available labels be used? We explore the flexibility of a newly-proposed semi-supervised approach: Fidelity-Weighted Learning. Specifically, we try to answer the question: Does the FWL teacher need to be a Gaussian Process?
1 Paulina Koziol, Magda Raczkowska, Danuta Liberda, Slawka Urbaniak, Czesława Paluszkiewicz, Wojciech Kwiatek, Tomasz Wrobel

Random forest classifier potential in the pancreatic cancer diagnosis with FT-IR imaging

Abstract: One of the greatest interests of the last few years is the implementation of machine learning (ML) methods into clinical tests. Exceptionally explored is the field of applying ML to spectroscopic imaging data of tissue specimen, in order to create histopathological models allowing future samples classification and diagnosis. Our main research is focused on using ML for creating full pancreatic cancer histopathology, based on chemical imaging in the form of Fourier Transform Infrared Spectroscopy.

2 Agnieszka Kraft, Ziad Al Bkhetan, Michał Kadlof, Dariusz Plewczynski

Inferring Three-Dimensional Chromatin Organization of Human Cells from Machine Learning Predicted Chromatin Interactions

Abstract: TBA

3 Piotr Woźnicki, Michał Kuźba, Piotr Migdał

Modelling response to trypophobia trigger using intermediate layers of ImageNet networks

Abstract: Trypophobia is a condition where a person experiences an intense and disproportionate fear towards clusters of small holes. In this paper, we approach the problem of detecting trypophobia triggers in images using Convolutional neural networks. We show that standard architectures such as VGG and ResNet50 can be successfully trained to recognise trypophobia patterns, achieving >94% accuracy on the task of binary classification. We also conduct experiments to analyze the nature of this phenomenon. We dissect the convolutional part of the network and train different models with reduced depth and number of parameters. We show that even extremely shallow networks are able to accurately label trypophobia images and correctly focus their attention on the trypophobia pattern as presented on the visual explanations. Therefore we suggest that patterns like trypophobia are of a primitive nature and can be quickly and effectively detected as such by human visual system. These triggers may automatically capture human attention and subsequently influence behavior.

4 Łukasz Mądry

Detecting out of distribution samples with random network distillation

Abstract: Detecting out of distribution samples is task of great importance in applications of machine learning in various contexts - e.g. medical imaging, self-driving cars or fraud detection. Being capable to assess whether given datapoint is close to our training set is crucial, since deep networks are known to output miscalibrated, too high probabilities. This means that we cannot rely on traditional measures of uncertainty, such as entropy or difference between first and second highest probability, since highest class is going to have probability close to one even for datapoints
far from training set. To solve this problem, we use method from called Random Network Distillation, described in part 2. In 3. we discuss both methods and results of our research.

5 Agnieszka Pocha, Krzysztof Hajto, Krzysztof Misztal

*Does FWL-teacher has to be a Gaussian Process?*

Abstract: Labels for biological datasets are expensive to obtain. One solution is to use models that do not require training (e.g., active contours are used for finding objects’ outline). However, these methods are unable to profit from available data and might require manual tuning for each new task. These obstacles can be overcome with deep learning models, though the question arises: How should the information from the little amount of available labels be used? We explore the flexibility of a newly-proposed semi-supervised approach: Fidelity-Weighted Learning. Specifically, we try to answer the question: Does the FWL teacher need to be a Gaussian Process?

6 Marcin Kowalik

*Estimation of Critical Temperature of high-temperature Superconductor by Neural Network*

Abstract: The value of critical temperature $T_c$ of high-temperature superconductor could be estimated from AC susceptibility vs temperature measurement using a pair of neural networks. In my poster I will describe the details of this approach and present a study of its performance and its limitations.

7 Krzysztof Rajda, Przemysław Spurek, Jacek Tabor

*GMM CWAE*

Abstract: One of the most important class of generative models is based on autoencoders architecture. Common autoencoder form latent space with some probabilistic distribution, mainly Gaussian. We aim to utilize advances of traditional clustering methods to improve autoencoders strength and flexibility by introducing Gaussian Mixture Model into its latent space. Here we present early results of our work.

8 Tomasz Danel

*Lichen Classifier Leveraging Recent Advances in Mobile Neural Networks*

Abstract: Recent studies on mobile neural network architectures show that it is possible to reduce the number of network parameters and mathematical operations without any significant drop of model accuracy. In order to achieve this, new microarchitectures are designed to replace standard convolutional layers. Novel optimization techniques, especially advances in weight compression and quantization, make it feasible to transfer deep architectures into mobile devices so that the inference phase of the model can be performed in real time – even with strict resource limitations. This work presents how these aforementioned methods can be combined to effectively create a mobile application which is built on top of a deep neural network. A lichen classification task was proposed to be the subject of this research. The result of the experiments is a highly optimized convolutional neural network inside an Android application that performs this task in real time.

9 Przemysław Spurek, Łukasz Maziarka, Aleksandra Nowak, Jacek Tabor, Stanisław Jastrzębski
Non-linear ICA based on Cramer-Wold metric

Abstract: Discovering an independent representation of signals is a useful tool with many downstream application. A popular algorithm for this task is the widely used Linear Independent Component Analysis (ICA). However, developing a nonlinear generalization of the ICA remains a challenging assignment. We ask whether a neural–based model optimizing a simple independence measure can perform a successful non–linear disentanglement. We build upon Adversarial Independent Component Analysis (ANICA) model and study the effect of using a closed–form optimization target instead of a discriminator–based independence measure.
1 Piotr Warchoł  
Jagiellonian University  

*Dynamical Isometry is Achieved in Residual Networks in a Universal Way for any Activation Function*

**Abstract:** We demonstrate that in residual neural networks (ResNets) dynamical isometry is achievable irrespectively of the activation function used. We do that by deriving, with the help of Free Probability and Random Matrix Theories, a universal formula for the spectral density of the input-output Jacobian at initialization, in the large network width and depth limit. The resulting singular value spectrum depends on a single parameter, which we calculate for a variety of popular activation functions, by analyzing the signal propagation in the artificial neural network. We corroborate our results with numerical simulations of both random matrices and ResNets applied to the CIFAR-10 classification problem. Moreover, we study the consequence of this universal behavior for the initial and late phases of the learning processes. We conclude by drawing attention to the simple fact, that initialization acts as a confounding factor between the choice of activation function and the rate of learning.

2 Michał Łukasik  
Google AI  

*Content Explorer: Recommending Novel Entities for a Document Writer*

**Abstract:** Background research is an essential part of document writing. Search engines are great for retrieving information once we know what to look for. However, the bigger challenge is often identifying topics for further research. Automated tools could help significantly in this discovery process and increase the productivity of the writer. In this paper, we formulate the problem of recommending topics to a writer. We consider this as a supervised learning problem and run a user study to validate this approach. We propose an evaluation metric and perform an empirical comparison of state-of-the-art models for extreme multi-label classification on a large data set. We demonstrate how a simple modification of the crossentropy loss function leads to improved results of the deep learning models.
1. Sabina Podlewska

Institute of Pharmacology of the Polish Academy of Sciences / Jagiellonian University Medical College

**Machine learning in the serve of searching for new drugs**

**Abstract:** Drug design is known for being an expensive and time-consuming process. Currently, computer-aided methods are an essential component in the drug discovery workflow, providing useful tools at its various stages. Their service starts at the very beginning, when computational tools are applied to identify new compounds with potentially desirable biological profiles. However, in silico evaluation is not only limited to the assessment of the activities of the investigated molecules towards considered targets but also involves analysis of their physicochemical and pharmacokinetic properties and potential toxicities.

The presentation will summarize the areas of application of computational methods in the search of new active compounds. It will focus on the application of machine learning methods, but also other molecular modeling methods will be presented. The examples of their usage in projects implemented in the Department of Medicinal Chemistry Institute of Pharmacology Polish Academy of Sciences will be provided.

2. Stanisław Jastrzębski, Maciej Szymczak, Agnieszka Pocha, Sabina Podlewska

**How to dock without docking - prediction of interaction fingerprints from 2D compound structure**

**Abstract:** A great variety of approaches is applied to find potential drug candidates – the most common practices include their de novo design or the selection on the basis of the evaluation of various compounds libraries, that is the so-called virtual screening (VS). One of the most popular methodology used in VS is docking. It enables the prediction of the compound orientation in the binding site and via the analysis of ligand-receptor interactions occurring in the obtained complex, the compound is scored in terms of its potential activity towards considered target.

However this process is computationally expensive and cannot be used for evaluation of very large compound libraries. In the study, we formulate methodology for predicting docking results (directly in the form of interaction fingerprints (IFPs) that are used for depiction of ligand-receptor complexes obtained in docking) using machine learning models. Preliminary results suggest that the proposed methodology achieves satisfactory accuracy.

3. Monika Piwowar, Jacek Dygut, Piotr Piwowar

**The GCA/ GCCA methods - a way of extracting genetic signatures from gene expression arrays**

**Abstract:** The authors present Grade Correspondence Analysis (GCA) and Grade Correspondence Cluster Analysis (GCCA) as a way to ordering and grouping of OMICS data sets. Based on transcriptomic data it was shown that the GCA method can be used to find regularities in the analysed collections and to create characteristic gene expression profiles for individual groups of patients. Based on the profiled gene expression of purified cell types isolated from 5
musculoskeletal tissues: tendon, bone, muscle, cartilage and ligament (GEO: GSE106292) and pre-treatment bone marrow from multiple myeloma patients (GEO: GSE2658) the GCA/GCCA methods is presented. The GCA and GCCA can be a method of choice for regularisation and grouping of data sets to get gene signatures. The result can become the starting point for further analysis, the characterisation of differentiating genes, biochemical processes and exploration of complex interactions.

4 Magdalena Wiercioch

3DST, A Novel Molecule Representation

**Abstract:** Molecular descriptors have been widely used to predict physicochemical properties and biological activities of molecules in order to identify promising drug leads from large chemical databases. Undoubtedly, machine learning methods offer an ancillary benefit to make molecule predictions more effective. However, the selection of representation is a challenge when choosing an algorithm. Nowadays, the available tools take into account the representations which do not cover sufficiently details connected with chemical data. What is more, the construction of good predictive models is difficult when the number of training samples is especially small. Herein, a novel descriptor that integrates spherical harmonics based descriptor and property characteristics is introduced. Performance of this approach was compared with several other methods including USR, MACCS and atom count representation. The talk will present the new achievements and put them into perspective.

5 Danuta Liberda, Michael Hermes, Paulina Kozioł, Magda Raczkowska, Nicholas Stone, Tomasz Wróbel

Impact of paraffin presence on Random Forest classification of FTIR hyperspectral images of esophagus cancer

**Abstract:** Nowadays Random Forest is one of the most frequently used classification method in FTIR (Fourier Transform Infrared Spectroscopy) chemical imaging of tissue samples. Random Forest has few advantages over other classification methods: it does not require large number of objects for model training, information about variable significance for model is available, optimization is quite simple and robustness for overfitting which can have a crucial impact on model prediction. In the case of tumor or healthy tissue classification it can consequently affect the decision about patient treatment. Data obtained with FTIR chemical imaging constitute hyperspectral images in which for each pixel spectrum with bands representing chemical composition of sample is obtained. One of the sample preparation approaches establishes sample measurement with paraffin wax, the other assumes sample dewaxing before measurement. Presence of paraffin wax in the sample results in a change of sample optical properties which in turn affects overall sample spectrum, including chemical composition changes. Spectral regions (variables) which comes from paraffin can be omitted in classification but there is a possibility that significant information is lost. Therefore, the goal of this study was to compare prediction power of two classification models constructed for FTIR chemical images of esophagus cancer and healthy tissue measured after application of two different sample preparation treatments.

6 Mateusz Półtorak, Mikołaj Morzy

Emotion Recognition in Raw Speech Recordings Using Convolutional Neural Networks and Spectrograms

**Abstract:** Emotion recognition from raw speech is an important and challenging task which has not found a satisfactory solution as of today. Over the last years, many SER (speech emotion
recognition) methods have been proposed. The majority of proposals depend on hand-crafted features based on the continuous and spectral characteristics of the audio signal. Sometimes, these methods are augmented with affective evaluation of spoken contents, which requires an automatic speech recognition system (ASR) output. The common weakness of current methods is the fact that their effectiveness depends directly on the set of selected features. In this work we present a method of automatic feature extraction for speech emotion recognition using convolutional neural networks (CNNs). We use spectrograms to perform convolutions and extract useful features from the raw audio signal. Experiments conducted on the IEMOCAP benchmark dataset prove the efficiency of our approach.
1. Tomasz Trzciński, Maciej Zięba
   Tooploox / Warsaw University of Technology
   
   BinGAN: Learning Compact Binary Descriptors with a Regularized GAN
   
   Abstract: In this talk, we present a novel regularization method for Generative Adversarial Networks, which allows the model to learn discriminative yet compact binary representations of image patches (image descriptors). We employ the dimensionality reduction that takes place in the intermediate layers of the discriminator network and train binarized low-dimensional representation of the penultimate layer to mimic the distribution of the higher-dimensional preceding layers. To achieve this, we introduce two loss terms that aim at: (i) reducing the correlation between the dimensions of the binarized low-dimensional representation of the penultimate layer i.e. maximizing joint entropy and (ii) propagating the relations between the dimensions in the high-dimensional space to the low-dimensional space. We present the evaluation of the resulting binary image descriptors on two challenging applications, image matching and retrieval, and achieve state-of-the-art results.

2. Marcin Możejko, Paweł Gora, Maciej Brzeski, Łukasz Mądry, Łukasz Skowronek
   
   Traffic signal settings optimization using gradient descent
   
   Abstract: We investigate performance of a gradient descent optimization applied to the traffic signal setting problem and compare it to genetic algorithms. We used neural networks as metamodels evaluating quality of signal settings and discovered that both optimization methods produce similar results, e.g., in both cases the accuracy of neural networks close to local optima depend on an activation function (e.g., TANH activation makes optimization process to converge to different minimas than ReLU activation).

3. Michał Zając, Konrad Żołna, Negar Rostamzadeh, Pedro Pinheiro
   
   Adversarial Framing for Image and Video Classification
   
   Abstract: Neural networks are prone to adversarial attacks. In general, such attacks deteriorate the quality of the input by either slightly modifying most of its pixels, or by occluding it with a patch. In this paper, we propose a method that keeps the image unchanged and only adds an adversarial framing on the border of the image. We show empirically that our method is able to successfully attack state-of-the-art methods on both image and video classification problems. Notably, the proposed method results in a universal attack which is very fast at test time.
4  Marcin Możejko, Michał Zmysłowski  
Penalizing approximated Hessian norm against adversarial attacks  

**Abstract:** We propose a novel method for tackling the problem of adversarial attacks. It is based on penalizing the Hessian Frobenius norm of a model output with respect to an input during training. We leverage the speed of approximating the norm by means of the difference between Jacobian-vector product of real and corrupted data during training. We test our method on different datasets and architectures. We achieve a significant decrease in number of adversarial examples compared to a situation when no regularisation is used or the sole Jacobian is being regularised.

5  Michał Sadowski, Aleksandra Grzegorczyk  
Image Inpainting with Gradient Attention  

**Abstract:** We present a novel modification of context encoder loss function, which results in more accurate and plausible inpainting. For this purpose, we introduce gradient attention loss component of loss function, to suppress the common problem of inconsistency in shapes and edges between the inpainted region and its context. To this end, the mean absolute error is computed not only for the input and output images, but also for their derivatives. Therefore, model concentrates on areas with larger gradient, which are crucial for accurate reconstruction. The positive effects on inpainting results are observed both for fully-connected and fully-convolutional models tested on MNIST and CelebA datasets.

6  Krzysztof Misztal, Przemysław Spurek, Jacek Tabor  
Image Stitching Based on Entropy Minimization  

**Abstract:** In this paper we present a method with closed analytic formula of stitching aligned images. It is obtained by choosing a statistically optimal global color change of one part of image. This approach, due to its numerical efficiency, is especially well-suited for merging big amount of satellite images into a single map. Moreover, we present solution of a general problem, how to find an optimal shift by $v$ of data $Y$ with respect to $v \in V$, so that the dataset $X, Y + v$ is maximally statistically consistent. We show that the solution is given in a closed analytic form.
1 Jarek Duda
Jagiellonian University

Hierarchical correlation reconstruction - between statistics and ML

Abstract: While machine learning techniques are very powerful, they have some weaknesses, like iterative optimization with many local minimums, large freedom of parameters, lack of their interpretability and accuracy control. From the other side we have classical statistics based on moments not having these issues, but providing only a rough description. I will talk about approach which combines their advantages: with MSE-optimal moment-like coefficients, but designed such that we can directly translate them into probability density. For multivariate case such basis of mixed moments asymptotically allows to accurately reconstruct any joint distribution, each coefficient can be independently and cheaply estimated, has a clear interpretation, and we have some control of its accuracy. I will also present its two applications: systematic enhancement of ARMA/ARCH-like modeling for any mixed moments and non-stationary time series (https://arxiv.org/pdf/1807.04119), and for credibility evaluation of income data: modeling continuous conditional probability distribution from a large number of variables of various types (https://arxiv.org/abs/1812.08040).

2 Marcin Mazur, Piotr Kościelniak

On some goodness of fit tests for normality based on the optimal transport distance

Abstract: We apply the optimal transport distance to construct two goodness of fit tests for (univariate) normality. The derived statistics are then compared with those used by the Shapiro-Wilk, the Anderson-Darling and the Cramer von Mises tests. In particular, we perform Monte Carlo experiments, involving computations of the test power against some selected alternatives and wide range of sample sizes, which show efficiency of the obtained test procedures.

3 Jan Kocoń

Improved recognition and normalisation of Polish temporal expressions using Cascade of Partial Rules

Abstract: This article introduces the issue of recognising temporal expressions for Polish language. We described what is temporal information and we presented TimeML specification, adapted to Polish as a model for the description of temporal expressions. The key aspect of the work is the improved method for the normalisation of temporal expressions, built on the top of the Conditional Random Fields model for the recognition of temporal expressions boundaries. We presented the experiments and conclusions drawn from them.

4 Jakub Chłędowski, Tomasz Wesołowski, Stanisław Jastrzębski

Improving Utilization of Lexical Knowledge in Natural Language Inference

Abstract: Natural language inference (NLI) is a central problem in natural language processing (NLP) of predicting the logical relationship between a pair of sentences. Lexical knowledge, which
represents relations between words, is often important for solving NLI problems. This knowledge can be accessed by using an external knowledge base (KB), but this is limited to when such a resource is accessible. Instead of using a KB, we propose a simple architectural change for attention based models. We show that by adding a skip connection from the input to the attention layer we can utilize better the lexical knowledge already present in the pretrained word embeddings. Finally, we demonstrate that our strategy allows to use an external source of knowledge in a straightforward manner by incorporating a second word embedding space in the model.

Igor Sieradzki, Igor Podolak

Distribution-Interpolation Trade off in Generative Models

Abstract: We investigate the properties of multidimensional probability distributions in the context of latent space prior distributions of implicit generative models. Our work revolves around the phenomena arising while decoding linear interpolations between two random latent vectors - regions of latent space in close proximity to the origin of the space are oversampled, which restricts the usability of linear interpolations as a tool to analyse the latent space. We show that the distribution mismatch can be eliminated completely by a proper choice of the latent probability distribution or using non-linear interpolations. We prove that there is a trade off between the interpolation being linear, and the latent distribution having even the most basic properties required for stable training, such as finite mean. We use the multidimensional Cauchy distribution as an example of the prior distribution, and also provide a general method of creating non-linear interpolations, that is easily applicable to a large family of commonly used latent distributions.