KIAS Center for AI and Natural Sciences 2024 Fall Workshop

November 4 (Mon.) - 8 (Fri.), 2024

Nest Hotel, Incheon

Speakers

Seungchan Ko (Inha University)

Dohyun Kwon (University of Seoul / KIAS)

Chieh-Hsin Lai (Sony AI)

Daniel D. Lee (Cornell University)

Jaeyong Lee (Chung-Ang University)

Hsuan-Tien Lin (National Taiwan University)

Guido Montúfar (UCLA / Max Planck Institute for Mathematics in the Sciences)

Krikamol Muandet (CISPA - Helmholtz Center for Information Security)

Jonggeol Na (Ewha Womans University)

Yung-Kyun Noh (Hanyang University / KIAS)

Kyungwoo Song (Yonsei University)

Masashi Sugiyama (The University of Tokyo / RIKEN-AIP)

Chulhee Yun (KAIST AI)

Organizers

Geonho Hwang (KIAS)

Changbong Hyeon (KIAS)

Dohyun Kwon (University of Seoul / KIAS)

Yung-Kyun Noh (Hanyang University / KIAS)

Program Schedule

	Nov. 4 (Mon)	Nov. 5 (Tue)		Nov. 6 (Wed)		Nov. 7 (Thu)			Nov. 8 (Fri)	
8:55 - 9:00		Opening Remarks								
9:00 – 10:00		Chair: Yung-Kyun Noh	Tutorial- Krikamol Muandet	Chair: Daniel Lee	Tutorial- Yung-Kyun Noh	Chair: Dohyun Kwon	Tutorial- Masashi Sugiyama	9:00 –9:40	Chair: Jaewoong Choi	Chulhee Yun
10:00 - 10:15			Break		Break		Break	9:40 -10:20		Seungchan Ko
10:15 - 10:55			Daniel D. Lee		Hsuan-Tien Lin		Guido Montúfar	10:20 -10:35		Break
10:55 - 11:15			Thomas Dieter Flacke		Ji Woong Yu		Daeseong Yong	10:35 - 11:15		Kyungwoo Song
11:15 - 11:30			Break		Break		Break	11:15 - 11:35		Sangwoong Yoon
11:30 - 12:10			Jonggeol Na		Chieh-Hsin Lai		Masashi Sugiyama	11:35 - 11:55		Break
								11:55 - 12:35		Dohyun Kwon
12:10 - 14:00		Lunch		Lunch		Lunch		12:35 – 14:00	Lunch	
14:00 - 14:20		Chair: Jaesung Choi	Jakwang Kim	Chair: Sungyoon Lee	Yeachan Park	Excursion (Muuido Island, Hanagae Beach)		14:00~	Closing	
14:20 - 14:40			Jaewoong Choi		Jaeyong Lee					
14:40 - 15:00			Break							
15:00 - 15:20	Check in		Krikamol Muandet		Break					
15:20 - 15:40					Jun Sur Park					
15:40 - 16:00		Poster session			Changhoon Song					
16:00 - 16:20					Kwang Hyun Cho					
16:20 - 16:40					Break					
16:40 - 17:20					Yung-Kyun Noh					
17:20 - 18:00										
18:00 – 20:00	Dinner	Dinner		Banquet		Dinner				

Nov. 5 (Tue)

8:55 - 9:00	Opening Remarks	
9:00 – 10:00	Tutorial- Krikamol Muandet	Causal Machine Learning
10:00 - 10:15	Break	
10:15 - 10:55	Daniel D. Lee	Fluid Dynamic Models in Machine Learning
10:55 - 11:15	Thomas Dieter Flacke	Color singlet, sextet or octet? Identifying and distinguishing new physics signals at colliders with machine learning
11:15 - 11:30	Break	
11:30 - 12:10	Jonggeol Na	Autonomous design of products and processes: Generative Chemical Transformer to reinforcement learning-guided combinatorial chemistry
12:10 - 14:00	Lunch	
14:00 - 14:20	Jakwang Kim	Adversarial robustness in classification via the lens of optimal transport
14:20 - 14:40	Jaewoong Choi	Improving Neural Optimal Transport via Displacement Interpolation
14:40 - 15:00	Break	
15:00 - 15:40	Krikamol Muandet	On Imprecise Generalisation: From Invariance to Heterogeneity
15:40 - 18:00	Poster session	
18:00 – 20:00	Dinner	

Nov. 6 (Wed)

9:00 – 10:00	Tutorial- Yung-Kyun Noh	Advances in Plug-in and Non-Plug-in Nearest Neighbor Methods for Utilizing Density Functionals
10:00 - 10:15	Break	
10:15 - 10:55	Hsuan-Tien Lin	Building Operational AI Weather Service in Taiwan
10:55 - 11:15	Ji Woong Yu	Recent Application of Machine Learning Force Field: Beyond Simple Molecular Dynamics Simulation
11:15 - 11:30	Break	
11:30 - 12:10	Chieh-Hsin Lai	Enhancing Accuracy and Efficiency in Diffusion Models
12:10 - 14:00	Lunch	
14:00 - 14:20	Yeachan Park	Understanding and acceleration of grokking phenomena in learning arithmetic operations via Kolmogorov-Arnold representation
14:20 - 15:00	Jaeyong Lee	Real-Time Solutions to PDEs with Neural Operators in Scientific Machine Learning
15:00 - 15:20	Break	
15:20 - 15:40	Jun Sur Park	WGFINNs: Weak-form Generic formalism informed neural networks
15:40 - 16:00	Changhoon Song	How does PDE order affect the convergence of PINNs?
16:00 - 16:20	Kwang Hyun Cho	Automized dataset collection for PES interpolation for pigment-protein complexes
16:20 - 16:40	Break	
16:40 - 17:20	Yung-Kyun Noh	Nonparametric f-Divergence Estimation and its Application to Eliminating Harmful Variables
17:20 - 18:00		
18:00 – 20:00	Banquet	

Nov. 7 (Thur)

9:00 – 10:00	Tutorial- Masashi Sugiyama	Machine Learning from Weak Supervision: An Empirical Risk Minimization Approach
10:00 - 10:15	Break	
10:15 - 10:55	Guido Montúfar	Implicit Bias of Mirror Descent for Shallow Neural Networks in Univariate Regression
10:55 - 11:15	Daeseong Yong	Dynamic Programming for Chain Propagator Computations in Polymer Field Theory Simulations
11:15 - 11:30	Break	
11:30 - 12:10	Masashi Sugiyama	Machine Learning under Distribution Shifts
12:10 - 14:00	Lunch	
14:40 - 18:00	Excursion	
18:00 – 20:00	Dinner	

Nov. 8 (Fri)

9:00 – 9:40	Chulhee Yun	Provable Benefit of Cutout and CutMix for Feature Learning
9:40 - 10:20	Seungchan Ko	Finite Element Operator Network: Theory and Computation
10:20 - 10:35	Break	
10:35 - 11:15	Kyungwoo Song	Robust Machine Learning via Sufficient Invariant and Causal-Aware In-Context Learning
11:15 - 11:35	Sangwoong Yoon	Value Gradient Sampling
11:35 - 11:55	Break	
11:55 - 12:35	Dohyun Kwon	A Fully First-Order Method for Stochastic Bilevel Optimization
12:35 - 14:00	Lunch	
14:00 -	Closing	

No.1

Causal Machine Learning

Krikamol Muandet

CISPA Helmholtz Center for Information Security

Society consists of diverse individuals, demographic groups, and institutions. Developing and deploying algorithmic models across these varied environments involves navigating numerous trade-offs. To create reliable machine learning algorithms capable of effective real-world interaction, addressing this heterogeneity is essential. In particular, the ability to identify and leverage causal relationships is a critical component of building reliable AI systems. In this tutorial, I will introduce the fundamentals of causal inference within the context of machine learning. I will begin by highlighting the mutual importance of causality in machine learning. Next, I will discuss how instrumental variables (IVs) can be used to mitigate the impact of unobserved confounders, enhancing the credibility of algorithmic decision-making and the reliability of models built from observational and heterogeneous data. Specifically, I will demonstrate how we can use machine learning tools, such as kernel methods and deep learning, to address potentially ill-posed problems in non-linear IV regression and proxy variable applications. Finally, I will emphasize that understanding how data is generated and how models can influence it is crucial for reliable human-machine interactions, especially when complete information about the data may be inaccessible

No.2

Fluid Dynamic Models in Machine Learning

Daniel D. Lee

Cornell University and KIAS

Fluid dynamic models play an important role in understanding the dynamics of recent machine learning models including reverse diffusion models and feature learning via stochastic gradient descent. I will describe connections between incompressible fluid models and probabilistic models, highlighting the role of conservation laws in both contexts. Specifically, the Fokker-Planck equation provides a framework to describe how velocity flow fields and diffusive processes influence the spatial and temporal evolution of the probability distribution. I will also discuss how these concepts apply to modeling the weights and kernel integral operators in neural networks during feature learning.

No.3

Color singlet, sextet or octet?

Identifying and distinguishing new physics signals at colliders with machine learning.

Thomas Dieter Flacke

KIAS

Machine learning provides a powerful tool for identifying and distinguishing new physics signals in large data samples with a lot of Standard Model background events at particle colliders. In this talk, I discuss the discovery and exclusion potential of the LHC with 3000 fb^-1 for pair- and single produced color singlet, sextet and octet spin 1 states in the 4 top final state. We implement a convolutional neural network combined with a fully connected DNN and compare its performance to a swin transformer network in tasks of a) identifying signal events in the event sample, b) determining the mass of the BSM spin 1 state, and c) discriminating color singlets, sextets and octets.

No.4

Autonomous design of products and processes: Generative Chemical Transformer to reinforcement learning-guided combinatorial chemistry

Jonggeol Na

Department of Chemical Engineering & Materials Science, Ewha Womans University

Recently, the concept of "autonomous design" has emerged as a way to optimize and design chemical products and processes autonomously. In the field of molecular design, various research achievements have been made from the perspective of inverse design to find molecular structures with desired properties. In this talk, I will introduce the Generative Chemical Transformer (GCT) (*Journal of Chemical Information and Modelling* 61 (12), 5804-5814, Front Cover) that enables high-speed inference while simultaneously matching multiple molecular target properties. At the same time, I mathematically show that extrapolation, a limitation of generative AI, is the same as the problem of predicting extreme properties of molecules. We discuss the methodology and results of reinforcement learning-guided combinatorial chemistry (RL-CC) (*Chemical Science*, 2024, 15, 7908-7925, Front Cover), which was developed to address this problem. Finally, I present the recent applications of RL-CC to (1) non-PFAS surfactant design, (2) amine design for reactive capture of carbon dioxide, and (3) OLED materials. In particular, the section on OLED material design introduces a new evaluation methodology based on bond dissociation energy. Our group utilizes self-supervised learning to efficiently predict the dissociation energy of large molecules at the OLED size level, resulting in improved energy predictions compared to existing methodologies. The combination of RL-CC and representation learning based on pre-trained large chemical models is expected to help solve a variety of problems in autonomous design, such as exploring unknown materials and finding optimal reaction conditions to produce them.

No.5

Adversarial robustness in classification via the lens of optimal transport

Jakwang Kim

PIMS, Mathematics of University of British Columbia

In this talk, I introduce the recent advance of the adversarial training problems of classification via optimal transport perspective. Since neural networks revolutionized the machine learning community, there have been tons of research to understand these objects. One critical issue is their instability against well-designed perturbation, which potentially causes serious problems in the application of deep learning. For this reason, people introduce so-called the adversarial training model for achieving more stable (robust) classifiers. Despite its practical importance, there has been no rigorous framework to describe and understand this model even until recently. In series papers, with Nicolas Garcia Trillos, Matt Jacobs and Matt Werenski, we do the following: (1) to connect the multiclass adversarial training problem to optimal transport and generalized barycenter problem, which first illustrates the geometry of this problem, (2) to prove the existence of robust classifiers, and unify variants of adversarial training models, and (3) to propose an efficient numerical scheme based on the combination of our theory and entropic optimal transport from computational optimal transport.

No.6

Improving Neural Optimal Transport via Displacement Interpolation

Jaewoong Choi

Korea Institute for Advanced Study

Optimal Transport (OT) theory investigates the cost-minimizing transport map that moves a source distribution to a target distribution. Recently, several approaches have emerged for learning the optimal transport map for a given cost function using neural networks. We refer to these approaches as the OT Map. OT Map provides a powerful tool for diverse machine learning tasks, such as generative modeling and unpaired image-to-image translation. However, existing methods that utilize max-min optimization often experience training instability and sensitivity to hyperparameters. In this paper, we propose a novel method to improve stability and achieve a better approximation of the OT Map by exploiting displacement interpolation, dubbed Displacement Interpolation Optimal Transport Model (DIOTM). We derive the dual formulation of displacement interpolation at specific time t and prove how these dual problems are related across time. This result allows us to utilize the entire trajectory of displacement interpolation in learning the OT Map. Our method improves the training stability and achieves superior results in estimating optimal transport maps. We demonstrate that DIOTM outperforms existing OT-based models on image-to-image translation tasks.

No.7

On Imprecise Generalisation: From Invariance to Heterogeneity

Krikamol Muandet

CISPA Helmholtz Center for Information Security

The ability to generalise knowledge across diverse environments stands as a fundamental aspect of both biological and artificial intelligence (AI). In recent years, significant advancements have been made in out-of-domain (OOD) generalisation, including the development of new algorithmic tools, theoretical advancements, and the creation of large-scale benchmark datasets. However, unlike in-domain (IID) generalisation, OOD generalisation lacks a precise definition, leading to ambiguity in learning objectives.

In this talk, I aim to clarify this ambiguity by arguing that OOD generalisation is challenging because it involves not only learning from empirical data but also deciding among various notions of generalisation. The intersection of learning and decision-making poses new challenges in modern machine learning, where distinct roles exist between machine learners (e.g., ML engineers) and model operators (e.g.,doctors).

To address these challenges, I will introduce the concept of imprecise learning, drawing connections to imprecise probability, and discuss our ICML 2024 paper (https://arxiv.org/abs/2404.04669) in the context of domain generalisation (DG) problems. By exploring the synergy between learning algorithms and decision-making processes, this talk aims to shed light on the complexities of OOD generalisation and pave the way for future advancements in the field.

No.8

Advances in Plug-in and Non-Plug-in Nearest Neighbor Methods for Utilizing Density Functionals

Yung-Kyun Noh

Hanyang University / Korea Institute for Advanced Study

The nearest neighbor distances in data space can be used for estimating probability densities. P'lugging in a well-estimated probabilities can, in principle, generate various useful machine learning methods, such as Bayes classification and feature selection using information-theoretic measures. However, regardless of how well the method performs in practical, real-world applications, an underlying issue with this straightforward approach is the misconception that the plug-in methodology is universally applicable with asymptotic guarantees. In this tutorial, I will provide a foundational understanding for constructing non-plug-in methods, beginning with bias-variance behavior in high-dimensional space, clarifying which aspect should be prioritized for reduction, and designing algorithms that use nearest neighbor information when plug-in methods are invalid for certain problems, such as *f*-divergence estimation. Historically, our estimators are connected to the analysis of nearest neighbors established by T. Cover in the 1960s, who have demonstrated potential in both theoretical and practical realms of machine learning, particularly in addressing the asymptotic behaviors of classification errors and density estimations.

No.9

Building Operational AI Weather Service in Taiwan

Hsuan-Tien Lin

National Taiwan University

Weather prediction stands as one of the most vital applications of big-data-driven artificial intelligence, profoundly affecting our daily lives. Despite the vast potential for advancements from the machine learning and computer vision communities, few researchers have had the unique opportunity to work closely with meteorologists. This talk will draw from the speaker's own collaborative experiences, illuminating both the successes and challenges faced, with the goal of inspiring and guiding future partnerships in weather and climate research. Additionally, the speaker will share insights into the intersections between meteorological applications and current developments in foundational models and generative artificial intelligence.

No.10

Recent Application of Machine Learning Force Field: Beyond Simple Molecular Dynamics Simulation

Ji Woong Yu

KIAS

Molecular Dynamics (MD) is a computational scientific tool that tracks the movement of atoms. The underlying physical interactions are obtained from force fields, which are sets of functional forms and parameters representing physical interactions between atoms. However, as evidenced by the existence of multiple force fields, determining the appropriate functional forms and parameters is non-trivial. Machine learning force fields (MLFFs) are an approach that delegates this task to neural networks. Since the early development of MLFFs in the late 2000s, there has been explosive growth in their development and use across physics, chemistry, and engineering. As demonstrated in extensive validation studies ranging from simple liquids and molecules to high-entropy alloys and biomolecules, MLFFs can serve as excellent substitutes for highly sophisticated first-principle calculations, such as density functional theory (DFT), at greatly reduced computational cost.

Recent interest has shifted towards exploring the discoveries and value MLFFs bring to atomic and molecular science beyond simple proof of concept. In this talk, I will start by briefly introducing popular MLFFs and some of the recent discoveries enabled by their use. Next, I will discuss the current limitations of MLFFs and how they can be circumvented through advanced molecular dynamics technique.

No.11

Enhancing Accuracy and Efficiency in Diffusion Models

Chieh-Hsin LAI

Sony AI

Diffusion models are powerful generative tools for high-fidelity data generation across various applications. However, they encounter two main challenges: slow sampling speed and fixed dimensionality (resolution). In this talk, I will introduce approaches to improve efficiency and accuracy in training and sampling by leveraging a unified concept of "consistency" rooted in the mathematical structure of diffusion models.

To accelerate sampling, I will introduce the Consistency Trajectory Model (CTM), which condenses a pre-trained diffusion model into a single neural network. CTM produces scores (log-density gradients) in one forward pass, enabling seamless traversal between any initial and final time along the Probability Flow ODE. This capability allows for new deterministic and stochastic sampling methods, including long jumps along ODE trajectories.

To address dimensionality limitations, I will present Progressive Growing of Diffusion Autoencoder (PaGoDA), which enhances the generator's resolution beyond that of the pre-trained diffusion model. By using a pre-trained low-resolution diffusion model to encode high-resolution data into a structured latent space, PaGoDA progressively increases the decoder's resolution without needing to retrain models during upsampling, thereby improving efficiency.

I hope my talk will stimulate collaborations and discussions across disciplines and contribute to AI-based Natural Sciences research.

No.12

Understanding and acceleration of grokking phenomena in learning arithmetic operations via Kolmogorov-Arnold representation

Yeachan Park

Center for AI and Natural Sciences, KIAS

We propose novel methodologies aimed at accelerating the grokking phenomenon, which refers to the rapid increment of test accuracy after a long period of overfitting as reported by Power et al. (2022). Focusing on the grokking phenomenon that arises in learning arithmetic binary operations via the transformer model, we begin with a discussion on data augmentation in the case of commutative binary operations. To further accelerate, we elucidate arithmetic operations through the lens of the Kolmogorov-Arnold (KA) representation theorem, revealing its correspondence to the transformer architecture: embedding, decoder block, and classifier. Observing the shared structure between KA representations associated with binary operations, we suggest various transfer learning mechanisms that expedite grokking. This interpretation is substantiated through a series of rigorous experiments. In addition, our approach is successful in learning two nonstandard arithmetic tasks: composition of operations and a system of equations. Furthermore, we reveal that the model is capable of learning arithmetic operations using a limited number of tokens under embedding transfer, which is supported by a set of experiments as well.

[1] Power, Alethea, et al. "Grokking: Generalization beyond overfitting on small algorithmic datasets." arXiv preprint arXiv:2201.02177 (2022).

No.13

Title: Real-Time Solutions to PDEs with Neural Operators in Scientific Machine Learning

Jaeyong Lee

Department of AI, Chung-Ang University

Recent advancements in deep learning have led to a surge in research focused on solving scientific problems under the "AI for Science." Among these efforts, Scientific Machine Learning (SciML) aims to address domain-specific data challenges and extract insights from scientific datasets through innovative methodological solutions. A particularly active area within SciML involves using neural operators to find real-time solutions to Partial Differential Equations (PDEs) as their parameters change. This presentation will discuss my latest research in this field, highlighting the significant potential of neural operators for solving complex physical phenomena.

No.14

WGFINNs: Weak-form Generic formalism informed neural networks

Jun Sur Park

Center for AI and Natural Sciences, KIAS

Numerous data-driven modeling studies have shown that employing a weak formulation of model equations with carefully selected test functions enhances robustness against noise. Inthis paper, we introduce the weak-form GENERIC formalism informed neural networks (WGFINNs)to improve the performance of the GENERIC formalism informed neural networks (GFINNs) for dis-covering underlying dynamics from noisy measurement data. Numerical examples demonstrate that,by leveraging the weak form, WGFINNs provide greater resilience to noise compared to GFINNs, enhancing the accuracy of the data-driven discovery of dynamics.

No.15

How does PDE order affect the convergence of PINNs?

Changhoon Song

KAIST

This talk addresses the inverse relationship between the order of partial differential equations (PDEs) and the convergence of gradient descent in physics-informed neural networks (PINNs) utilizing Rectified Power Unit (RePU) activation. By integrating PDE constraints into the loss function, PINNs inherently require the computation of derivatives up to the order of the PDE. While empirical observations suggest that PINNs often struggle with convergence for high-order or high-dimensional PDEs, a thorough theoretical explanation has been lacking. This talk provides a theoretical foundation for these challenges, showing that the gradient flow is less likely to converge as the PDE order increases. Additionally, we explore the impact of dimensionality on convergence, which is further compounded by higher PDE orders. To mitigate these issues, we propose a variable splitting technique that decomposes high-order PDEs into a system of lower-order PDEs. We demonstrate that this approach improves the likelihood of convergence to the global optimum. Numerical experiments are presented to support our theoretical findings and highlight the practical implications of our method.

No.16

Automized dataset collection for PES interpolation for pigment-protein complexes

Kwang Hyun Cho

Center for AI and Natural Sciences, KIAS

khcho@kias.re.kr

Nature has evolved various species to capture light energy and convert it into chemical energy during photosynthesis. Light harvesting complexes, composed of spatially arranged pigments and surrounding proteins, play an essential role in this process by delivering light energy through excitation energy transfer (EET) process. The high efficiency of the EET process is a key factor in the overall energy yield of photosynthesis. Understanding the molecular origins of this efficient process requires a detailed investigation of how molecular motions affect pigment quantum states, which can be achieved through molecular dynamics simulations. A key challenge in these simulations is providing an accurate potential energy surface (PES) while maintaining computational efficiency. Previously, we have demonstrated that an interpolation-based approach, with properly sampled datasets, can capture the realistic dynamics and environmental noise of the pigment molecules in light-harvesting complexes, such as the FMO and LH2 complex. However, more systematic approaches are still required, as any other complex requires its own PES due to its unique protein environment. Here, I propose utilizing machine learning techniques to automize dataset collection by reusing pre-obtained data for PES interpolation. This approach aims to systematically construct PESs with minimal effort and to allow for the comparisons of the molecular environments across various pigment-protein complexes. It can provide insights into the fundamental molecular mechanisms underlying efficient energy transfer, which may be ubiquitous in nature.

No.17

Nonparametric *f*-Divergence Estimation and its Application to Eliminating Harmful Variables

Yung-Kyun Noh

Hanyang University / Korea Institute for Advanced Study

Nearest neighbor methods are well-regarded for their simplicity and scalability, allowing parallel computation without extensive implementation effort. This research explores advancements in nearest neighbor methods tailored for *f*-divergence estimation and their applications in adjusting deep learning models for trustworthiness. I will introduce a systematic non-plug-in method using *k*-nearest neighbors to construct a nonparametric estimator for a target *f*-divergence. The proposed method leverages the inverse Laplace transform, offering a contrast to previous plug-in methodologies, which have theoretical shortcomings when using a fixed *k*. Applications of these methods will be briefly discussed to address various challenges confronted in artificial intelligence, such as handling imperfect information, ensuring fairness, and eliminating artifacts in simulated data.

No.18

Machine Learning from Weak Supervision: An Empirical Risk Minimization Approach

Masashi Sugiyama

RIKEN and The University of Tokyo, Japan.

https://www.ms.k.u-tokyo.ac.jp/sugi/ *sugi@k.u-tokyo.ac.jp*

Machine learning (ML) is a sub-field of artificial intelligence (AI), which is aimed at investigating computer algorithms that improve themselves automatically through experience. ML has been one of the most evolved and deepened research topics in science and technology in the early 21st century, and it is the technology that boosted the use of AI in the real world. Nowadays, ML-based AI systems have been deployed vitally in pioneering new business as well as advancing scientific research and technological development.

So far, success of ML has mainly been in the virtual world, such as e-commerce, social networks, and gaming, since ML-based AI systems need to be trained with big data that contain rich supervised information. However, once we try to apply ML to physical-world problems in our real world, such as medical diagnosis, natural disaster, and education, it is extremely difficult or even impossible to collect such a huge amount of fully supervised data. Thus, data collection is one of the critical bottlenecks for AI to be further penetrated in our society. Thus, there is an urgent need to develop novel theory and algorithm of ML that allow us to train ML-based AI systems from limited supervision.

This lecture is aimed at providing basics and practical algorithms of weakly supervised classification, based on the monograph we published in 2022 [1]. By weakly supervised classification, we do not mean that we try to train a classifier from small training data, which is mathematically not possible without imposing strong assumptions on data and models. Instead, we try to train a classifier with a large amount of data that can be easily collected. Such easily collectible data usually contain weaker supervised information than expensive fully supervised data, but we will show that it is possible to train a classifier only from such weakly supervised data as if we have fully supervised data.

[1] Sugiyama, M., Bao, H., Ishida, T., Lu, N., Sakai, T., & Niu, G. Machine Learning from Weak Supervision: An Empirical Risk Minimization Approach, MIT Press, Cambridge, Massachusetts, USA, 2022.

No.19

Implicit Bias of Mirror Descent for Shallow Neural Networks in Univariate Regression

Guido Montufar

University of California, Los Angeles

We examine the implicit bias of mirror flow in univariate least squares error regression with wide and shallow neural networks. For a broad class of potential functions, we show that mirror flow exhibits lazy training and has the same implicit bias as ordinary gradient flow when the network width tends to infinity. For ReLU networks, we characterize this bias through a variational problem in function space. Our analysis includes prior results for ordinary gradient flow as a special case and lifts limitations which required either an intractable adjustment of the training data or networks with skip connections. We further introduce scaled potentials and show that for these, mirror flow still exhibits lazy training but is not in the kernel regime. For networks with absolute value activations, we show that mirror flow with scaled potentials induces a rich class of biases, which generally cannot be captured by an RKHS norm. A takeaway is that whereas the parameter initialization determines how strongly the curvature of the learned function is penalized at different locations of the input space, the scaled potential determines how the different magnitudes of the curvature are penalized. This is work with Shuang Liang.

No.20

Dynamic Programming for Chain Propagator Computations in Polymer Field Theory Simulations

Daeseong Yong, KIAS

We present an algorithmic approach that optimizes chain propagator computations in polymer field theory simulations. These computations have recursive structures and there are heavily overlapping computations for branched polymers. By employing dynamic programming, these redundant computations are systematically avoided for any mixture of arbitrary acyclic branched block copolymers. We demonstrate that our approach achieves optimal time complexity for various polymeric systems, including multi-arm star-shaped polymers, comb polymers, dendrimers, and homopolymer mixtures. This work paves the way for the development of efficient open-source software and holds potential applications in automated searches for inverse design.

No.21

Machine Learning under Distribution Shifts

Masashi Sugiyama

RIKEN and The University of Tokyo, Japan.

https://www.ms.k.u-tokyo.ac.jp/sugi/ sugi@k.u-tokyo.ac.jp

A common assumption in standard machine learning methods is that the data used for training a predictor follow the same probability distribution as the data used for testing the prediction performance in the inference phase. However, in many real-world applications, this common assumption is often violated, e.g., due to changing environments over time or sample selection bias caused by privacy concerns. Such a situation is called distribution shift, and how to overcome the distribution shift is an urgent challenge in the machine learning community.

In this talk, I will first give an overview of the classical importance weighting approach to distribution shift adaptation, which consists of an importance estimation step and an importance-weighted training step [1,2]. Then, I will present a more recent approach that simultaneously estimates the importance weight and trains a predictor. I will also discuss a more practical scenario of sequential distribution shifts, where the data distributions change sequentially over time. Finally, I will discuss ongoing challenges such as joint distribution shift, out-of-distribution adaptation, and more.

[1] Sugiyama, M. & Kawanabe, M. Machine Learning in Non-Stationary Environments: Introduction to Covariate Shift Adaptation, MIT Press, Cambridge, Massachusetts, USA, 2012.

[2] Quiñonero-Candela, J., Sugiyama, M., Schwaighofer, A., & Lawrence, N. D. (Eds.), Dataset Shift in Machine Learning, MIT Press, Cambridge, Massachusetts, USA, 2009.

No.22

Provable Benefit of Cutout and CutMix for Feature Learning

Chulhee Yun

KAIST

Patch-level data augmentation techniques such as Cutout and CutMix have demonstrated significant efficacy in enhancing the performance of image-based tasks. However, a comprehensive theoretical understanding of these methods remains elusive. In this paper, we study two-layer neural networks trained using three distinct methods: vanilla training without augmentation, Cutout training, and CutMix training. Our analysis focuses on a feature-noise data model, which consists of several label-dependent features of varying rarity and label-independent noises of differing strengths. Our theorems demonstrate that Cutout training can learn low-frequency features that vanilla training cannot, while CutMix training can even learn rarer features that Cutout cannot capture. From this, we establish that CutMix yields the highest test accuracy among the three. Notably, our novel analysis reveals that CutMix training makes the network learn all features and noise vectors "evenly" regardless of the rarity and strength, which provides an interesting insight into understanding patch-level augmentation. This is joint work with Junsoo Oh (KAIST), to be presented at NeurIPS 2024 as a spotlight paper.

No.23

Finite Element Operator Network: Theory and Computation

Seungchan Ko

Department of Mathematics, Inha University

Partial differential equations (PDEs) underlie our understanding and prediction of natural phenomena across numerous fields, including physics, engineering, and finance. However, solving parametric PDEs is a complex task that necessitates efficient numerical methods. In this talk, I introduce a novel approach for solving parametric PDEs using a Finite Element Operator Network (FEONet). The proposed method leverages the power of deep learning in conjunction with traditional numerical methods, specifically the finite element method, to solve parametric PDEs in the absence of any paired input-output training data. I will demonstrate the effectiveness of our approach on several benchmark problems and show that it outperforms existing methods in terms of accuracy, generalization, and computational flexibility. Furthermore, I will also provide theoretical convergence analysis to support our approach in the numerical analysis framework.

No.24

Robust Machine Learning via Sufficient Invariant and Causal-Aware In-Context Learning

Taero Kim, Hoyoon Byun, Subeen Park, SungJun Lim, Gyeongdeok Seo, Taero Kim, Jihee Kim, **Kyungwoo Song**

Yonsei University

Machine learning models often underperform when faced with distribution shifts between training and testing data. To enhance robustness, we introduce two approaches. First, we propose the Sufficient Invariant Learning (SIL) framework, a novel method designed to capture a comprehensive set of invariant features across diverse environments. To support SIL, we develop Adaptive Sharpness-aware Group DRO (ASGDRO), an algorithm that identifies flat minima to ensure effective generalization across domains. Additionally, we construct a benchmark dataset specifically tailored to evaluate the sufficiency and diversity of invariant feature learning under various distribution shifts. Second, we present Causal-aware In-Context Learning (CCL), which utilizes VAE-based causal representation learning to select task-relevant examples for in-context learning. Unlike traditional methods that rely on superficial linguistic similarity, CCL focuses on causally relevant features, improving performance in out-of-distribution scenarios. Our theoretical and empirical results confirm that SIL and CCL significantly enhance model resilience to distributional changes, ensuring reliable performance across a wide range of challenging environments.

No.25

Value Gradient Sampling

Sangwoong Yoon

AI Research Fellow

Korea Institute for Advanced Study (KIAS)

We propose the Value Gradient Sampler (VGS), a trainable sampler inspired by optimal control. VGS generates samples from a given unnormalized density (i.e., energy) by drifting and diffusing randomly initialized particles, similar to Langevin Monte Carlo. Minimizing the KL divergence between the target density and the samples, VGS can synthesize accurate samples in fewer steps. We formulate this KL divergence minimization as an optimal control problem and apply value-based dynamic programming to obtain the optimal drift and diffusion at each sampling step. During sampling, particles drift along the gradient of the learned value function, which is learned using standard reinforcement learning techniques like temporal difference learning. Being a fast, adaptive, and accurate sampling method, VGS can be applied to generate negative samples in contrastive divergence training of energy-based models. We demonstrate the effectiveness of VGS in training energy-based models for industrial anomaly detection.

No.26

A Fully First-Order Method for Stochastic Bilevel Optimization

Dohyun Kwon

Department of Mathematics, University of Seoul / Center for AI and Natural Sciences, KIAS

The problem of stochastic bilevel optimization has been the focus of extensive study in recent years. Although many optimization methods have been proposed to address bilevel problems, existing approaches often require potentially expensive calculations involving the Hessians of lower-level objectives. The primary technical challenge lies in tracking the lower-level solutions as upper-level variables change. We introduce a Fully First-order Stochastic Approximation (F2SA) method, which only relies on first-order gradient oracles. Additionally, we analyze the complexity of first-order methods under minimal assumptions and provide matching lower bounds. This talk is based on joint work with Jeongyeol Kwon, Hanbaek Lyu, Stephen Wright, and Robert Nowak (UW-Madison).

Poster No.1

Bayesian active learning of molecular properties with small initial data

Minyoung Ha

Materials Research Center,

Samsung Advanced Institute of Technology (SAIT)

Designing novel materials with limited data remains a critical challenge in materials science. Since large-scale data is often unavailable in early-stage research, materials discovery typically relies on small, proof-of-concept datasets and a trial-and-error process guided by intuition. To address this, we propose a two-stage Bayesian active learning workflow for molecular property prediction (MPP) and design of experiments (DOE), working with an initial dataset as small as 50 points.

Our approach utilizes a message-passing graph neural network (GNN) to map molecular structures into fixed-length graph embeddings, followed by a Gaussian process regressor (GPR) to predict molecular properties. Extensive virtual experiments were conducted for various tasks across multiple molecular domains to compare the MPP and DOE performances of graph embedding methods and GPR kernel choices. We demonstrate that system-agnostic embeddings such as random graph embeddings provide a solid baseline for a wide range of problems, while graph embeddings from a pre-trained GNN significantly enhance performance if a large database of associated molecular properties is available. Interestingly, although MPP and DOE performances are generally correlated, strong cross-validation performance in MPP does not always translate to effective DOE. These findings suggest new strategies for data-driven and data-efficient materials discovery, especially in the early, data-scarce stages of research, where traditional approaches are less effective.

Poster No.2

Interpreting Bayesian Methods Through the Geometry of Parameter Spaces

Cheongjae Jang

Hanyang University

In naive maximum a posteriori (MAP) estimation, the mode of the posterior distribution may shift depending on the choice of coordinates in the parameter space. This result is undesirable as it suggests that estimation outcomes can be influenced by reparametrization, which is irrelevant for inference. This issue can be addressed by appropriately considering the geometry of the parameter space. In this presentation, we will interpret various Bayesian methods through the lens of parameter space geometry and discuss prospective directions for future research.

Poster No.3

DiffAlign: Flexible Molecular Alignment Using Diffusion

Iljung Kim

Department of Computer Science, Hanyang University, Seoul, Republic of Korea

Flexible molecular alignment is fundamental in computational chemistry and structural biology, serving as cornerstone for understanding molecular interactions and function. Recent methods approach the flexible molecular alignment problem as a score optimization task, leaving room for improvements through artificial intelligence. To address this, we introduce DiffAlign, a diffusion model that frames flexible molecular alignment as a generative modeling problem. DiffAlign is trained using an SE(3) equivariant conditional diffusion approach. By training the model within a conditional diffusion framework, we can leverage classifier-free guidance, enabling DiffAlign to align one molecule to another with high precision. DiffAlign achieves a root-mean-square deviation (RMSD) of 0.33 for self-alignment, significantly improving upon the 0.59 RMSD obtained by traditional state-of-the-art methods for DUD-E INHA dataset. Additionally, DiffAlign demonstrates a major efficiency advantage, solving the alignment problem in just 2.6 seconds on average, compared to 41 seconds for previous state-of-the-art methods.

Poster No.4

Quantifying Parameter Importance in Neural Networks: Beyond Shapley Value approach

Do-Hoon Kim

Department of Computer Science

Hanyang University, Korea

Traditional interpretation and explanation methods of neural network (NN) models have predominantly focused on feature importance, addressing the question "Which features are most important to the model predictions?" However, our primary interest in model interpretation lies in '**parameter importance**', seeking to answer, "Which parameters are most crucial to the model itself?" Regardless of how important feature information is, it cannot influence model predictions without parameters to convey that information. Assigning importance values to parameters, while considering their inherent characteristics, is crucial for accurate model optimization and effective model compression and fundamental for developing reliable and efficient machine learning systems. The Shapley value, originating from cooperative game theory in economics, has been adopted for parameter importance value due to its solid foundation in fairly distributing contributions. However, we argue that the Shapley value, which averages parameter contributions across all possible combinations of parameters, where only a portion of the parameters is utilized in the model, overlooks critical interactions and additional information within these combinations. This limitation is particularly evident in scenarios where parameter interactions lead to overlapping information or enable feature information flow within specific parameter combinations. We introduce a novel perspective that emphasizes the roles of individual parameters within the context of all possible parameter combinations in NNs. Our main contribution is the development of criteria that enable informed decisions about retaining or removing specific parameters based on their roles in NNs. Additionally, we offer a more comprehensive understanding of parameter significance, extending beyond the traditional approaches provided by the Shapley value.

Poster No.5

A Unified Confidence Sequence for Generalized Linear Models, with Applications to Bandits

Junghyun Lee[1], Se-Young Yun[1], Kwang-Sung Jun[2]

[1]*Kim Jaechul Graduate School of AI, KAIST*

[2]*Department of Computer Science, University of Arizona*

We present a unified likelihood ratio-based confidence sequence (CS) for *any* (self-concordant) generalized linear models (GLMs) that is guaranteed to be convex and numerically tight. We show that this is on par or improves upon known CSs for various GLMs, including Gaussian, Bernoulli, and Poisson. In particular, for the first time, our CS for Bernoulli has a $\mathrm{poly}(S)$-free radius where S is the norm of the unknown parameter. Our first technical novelty is its derivation, which utilizes a time-uniform PAC-Bayesian bound with a uniform prior/posterior, despite the latter being a rather unpopular choice for deriving CSs. As a direct application of our new CS, we propose a simple and natural optimistic algorithm called **OFUGLB** applicable to *any* generalized linear bandits (**GLB**; Filippi et al., 2010). Our analysis shows that the celebrated optimistic approach simultaneously attains state-of-the-art regrets for various self-concordant (not necessarily bounded) **GLB**s, and even $\mathrm{poly}(S)$-free for bounded **GLB**s, including logistic bandits. The regret analysis, our second technical novelty, follows from combining our new CS with a new proof technique that completely avoids the previously widely used self-concordant control lemma (Lemma 9, Faury et al., 2020). We numerically verify that **OFUGLB** significantly outperforms, or is at least at par with, prior algorithms for logistic and Poisson bandits.

Poster No.6

Eigenvalue-Based Preprocessing for Tissue Extraction from Pathology Image Slides

DoHyun Lim

Department of Computer Science

Hanyang University

The objective of this study is to effectively extract patches containing only tissue from pathology image slides. Pathology image slides include not only tissue but also unwanted elements such as background and markings, necessitating precise preprocessing for accurate tissue extraction. To address this, we devised a preprocessing method utilizing the eigenvalues and eigenvectors derived from the covariance matrix calculated from the image's RGB values.

We hypothesized that the brightness difference between tissue and background induces the greatest variability. Consequently, the direction with the largest variance corresponds to the first principal eigenvector, and its magnitude corresponds to the first eigenvalue. Visualization of the RGB pixel data in a 3D scatter plot confirmed a trend where brightness increases in the direction of the first eigenvector.

Notably, tissue patches exhibited significantly larger first eigenvalues compared to background patches, demonstrating effective separation between tissue and background. Additionally, we observed clustering of regions representing key features within the tissue in the 3D scatter plot, indicating successful extraction of data that reflects important tissue characteristics. This preprocessing method enabled accurate extraction of tissue patches.

This study validates the efficacy of the eigenvalue-based method in the preprocessing stage of pathology image analysis and is expected to lay the foundation for the development of a Foundation Model optimized for pathology in the future.

Poster No.7

Provable Benefit of Cutout and CutMix for Feature Learning

Junsoo Oh

KAIST AI

Patch-level data augmentation techniques such as Cutout and CutMix have demonstrated significant efficacy in enhancing the performance of image-based tasks. However, a comprehensive theoretical understanding of these methods remains elusive. In this paper, we study two-layer neural networks trained using three distinct methods: vanilla training without augmentation, Cutout training, and CutMix training. Our analysis focuses on a feature-noise data model, which consists of several label-dependent features of varying rarity and label-independent noises of differing strengths. Our theorems demonstrate that Cutout training can learn features with low frequencies that vanilla training cannot, while CutMix training can even learn rarer features that Cutout cannot capture. From this, we establish that CutMix yields the highest test accuracy among the three. Our novel analysis reveals that CutMix training makes the network learn all features and noise vectors ``evenly'' regardless of the rarity and strength, which provides an interesting insight into understanding patch-level augmentation.

Poster No.8

Machine Learning Force Fields for Ionic Liquids

Anseong Park, Won Bo Lee*

School of Chemical and Biological Engineering, Seoul National University, Republic of Korea,

wblee@snu.ac.kr

The development of machine learning force fields (MLFFs) offers a promising alternative, aiming to reproduce potential energy surfaces (PES) based on DFT data. However, the quality of MLFFs largely depends on the expertise of researchers in preparing training datasets and tuning hyperparameters. Unlike traditional FFs, which are systematically built and transferable, MLFFs may face uncertainties in covering rare events during simulations, especially in systems with diverse atomic types. In this study, DeePMD was used to construct MLFFs, and the results were compared to classic and polarizable FF MD simulations. The findings reveal that incorporating non-equilibrated (nEQ) datasets enhances MLFF performance, yet discrepancies with polarizable FF results raise questions about the correctness and completeness of MLFF simulations in capturing complex behaviors.

Poster No.9

DASH: Warm-Starting Neural Network Training Without Loss of Plasticity Under Stationarity

Baekrok Shin

KAIST AI

Warm-starting neural networks by initializing them with previously learned weights is appealing, as practical neural networks are often deployed under a continuous influx of new data. However, this often leads to a phenomenon known as *loss of plasticity*, where the network loses its ability to learn new information and thereby shows worse generalization performance than those trained from scratch. While this issue has been actively studied in non-stationary data distributions (e.g., in reinforcement learning), it surprisingly occurs even when the data distribution remains stationary, and its underlying mechanism is poorly understood. To address this gap, we develop a learning framework that emulates real-world neural network training scenarios. Under this framework, we identify noise memorization as the primary cause of the loss of plasticity when warm-starting the neural networks in stationary data distributions. Motivated by this discovery, we propose an effective method called **Direction-Aware SHrinking (DASH)** to mitigate the loss of plasticity. DASH aims to selectively forget previously memorized noise while aiming to preserve learned features, based on the loss gradient computed from newly introduced data. We validate our approach in vision tasks using diverse datasets, models, and optimizers, demonstrating consistent improvements in test accuracy and training efficiency.

Poster No.10

Machine learning integrated Fourier transform for enhanced signal analysis

Jihye Kim[1], Woo Seok Lee[2], Jaehyun Jung[3] and Taegeun Song[1]

[1]*Department of Data Information and Physics, Kongju National University, Korea*

[2] *1st Bioterapuritics, Korea*

[1] *Department of Applied Mathatics, Kongju National University, Korea*

We propose a novel approach for performing Fourier transforms using a single hidden-layer perceptron with a sine activation function. The proposed approach demonstrates remarkable capabilities, including the ability to identify frequency components even with incomplete cycle data and an effectively infinite observation time akin to the delta peak. Due to its mathematical equivalence to the Fourier transform, the proposed method benefits from being effective even when overfitting occurs, by taking advantage of its inherent properties to improve results. As a practical application, we have applied this method to a quantum interferometer and suggest its potential use for missing data prediction and super-resolution tasks.

Poster No.11

Towards Calibrated Robust Fine-Tuning of Vision-Language Models

Changdae Oh, Hyesu Lim, Mijoo Kim, Dongyoon Han, Sangdoo Yun, Jaegul Choo,

Alexander G Hauptmann, Zhi-Qi Cheng, **Kyungwoo Song**

University of Wisconsin–Madison, KAIST, Chung-Ang University, NAVER AI Lab, Carnegie Mellon University,

Yonsei University

Improving out-of-distribution (OOD) generalization through in-distribution (ID) adaptation is a primary goal of robust fine-tuning methods beyond the naive fine-tuning approach. However, despite decent OOD generalization performance from recent robust fine-tuning methods, OOD confidence calibration for reliable machine learning has not been fully addressed. This work proposes a robust fine-tuning method that improves both OOD accuracy and calibration error in Vision Language Models (VLMs). Firstly, we show that both types of errors have a shared upper bound consisting of two terms of ID data: 1) calibration error and 2) the smallest singular value of the input covariance matrix. Based on this insight, we design a novel framework that conducts fine-tuning with a constrained multimodal contrastive loss enforcing a larger smallest singular value, which is further aided by the self-distillation of a moving averaged model to achieve well-calibrated prediction. Starting from an empirical validation of our theoretical statements, we provide extensive experimental results on ImageNet distribution shift benchmarks that demonstrate the effectiveness of our method.