3801. Linear Bandits with Memory

链接：https://iclr.cc/virtual/2025/poster/31492 abstract： Nonstationary phenomena, such as satiation effects in recommendations, have mostly been modeled using bandits with finitely many arms. However, the richer action space provided by linear bandits is often preferred in practice. In this work, we introduce a novel nonstationary linear bandit model, where current rewards are influenced by the learner's past actions in a fixed-size window. Our model, which recovers stationary linear bandits as a special case, leverages two parameters: the window size $m \ge 0$, and an exponent $\gamma$ that captures the rotting ($\gamma < 0)$ or rising ($\gamma > 0$) nature of the phenomenon. When both $m$ and $\gamma$ are known, we propose and analyze a variant of OFUL which minimizes regret against cyclic policies. By choosing the cycle length so as to trade-off approximation and estimation errors, we then prove a bound of order $\sqrt{d}\,(m+1)^{\frac{1}{2}+\max{\gamma,0}}\,T^{3/4}$ (ignoring log factors) on the regret against the optimal sequence of actions, where $T$ is the horizon and $d$ is the dimension of the linear action space. Through a bandit model selection approach, our results are then extended to the case where both $m$ and $\gamma$ are unknown. Finally, we complement our theoretical results with experiments comparing our approach to natural baselines.

3802. Do Stochastic, Feel Noiseless: Stable Stochastic Optimization via a Double Momentum Mechanism

链接：https://iclr.cc/virtual/2025/poster/27684 abstract： Optimization methods are crucial to the success of machine learning, with Stochastic Gradient Descent (SGD) serving as a foundational algorithm for training models. However, SGD is often sensitive to the choice of the learning rate, which necessitates extensive hyperparameter tuning. In this work, we introduce a new variant of SGD that brings enhanced stability in two key aspects. First, our method allows the use of the same fixed learning rate to attain optimal convergence rates regardless of the noise magnitude, eliminating the need to adjust learning rates between noiseless and noisy settings. Second, our approach achieves these optimal rates over a wide range of learning rates, significantly reducing sensitivity compared to standard SGD, which requires precise learning rate selection.Our key innovation is a novel gradient estimator based on a double-momentum mechanism that combines two recent momentum-based techniques. Utilizing this estimator, we design both standard and accelerated algorithms that are robust to the choice of learning rate. Specifically, our methods attain optimal convergence rates in both noiseless and noisy stochastic convex optimization scenarios without the need for learning rate decay or fine-tuning. We also prove that our approach maintains optimal performance across a wide spectrum of learning rates, underscoring its stability and practicality. Empirical studies further validate the robustness and enhanced stability of our approach.

3803. Reexamining the Aleatoric and Epistemic Uncertainty Dichotomy

链接：https://iclr.cc/virtual/2025/poster/31334 abstract： When discussing uncertainty estimates for the safe deployment of AI agents in the real world, the field typically distinguishes between aleatoric and epistemic uncertainty. This dichotomy may seem intuitive and well-defined at first glance, but this blog post reviews examples, quantitative findings, and theoretical arguments that reveal that popular definitions of aleatoric and epistemic uncertainties directly contradict each other and are intertwined in fine nuances. We peek beyond the epistemic and aleatoric uncertainty dichotomy and reveal a spectrum of uncertainties that help solve practical tasks especially in the age of large language models.

3804. Generalizable Motion Planning via Operator Learning

链接：https://iclr.cc/virtual/2025/poster/29468 abstract： In this work, we introduce a planning neural operator (PNO) for predicting the value function of a motion planning problem. We recast value function approximation as learning a single operator from the cost function space to the value functionspace, which is defined by an Eikonal partial differential equation (PDE). Therefore, our PNO model, despite being trained with a finite number of samples at coarse resolution, inherits the zero-shot super-resolution property of neural operators. We demonstrate accurate value function approximation at 16× the training resolution on the MovingAI lab’s 2D city dataset, compare with state-of-the-art neural valuefunction predictors on 3D scenes from the iGibson building dataset and showcase optimal planning with 4-joint robotic manipulators. Lastly, we investigate employing the value function output of PNO as a heuristic function to accelerate motion planning. We show theoretically that the PNO heuristic is $\epsilon$-consistent by introducing an inductive bias layer that guarantees our value functions satisfy the triangle inequality. With our heuristic, we achieve a $30$% decrease in nodes visited while obtaining near optimal path lengths on the MovingAI lab 2D city dataset, compared to classical planning methods (A$^\ast$, RRT$^\ast$).

3805. ADAM: An Embodied Causal Agent in Open-World Environments

链接：https://iclr.cc/virtual/2025/poster/29794 abstract： In open-world environments like Minecraft, existing agents face challenges in continuously learning structured knowledge, particularly causality. These challenges stem from the opacity inherent in black-box models and an excessive reliance on prior knowledge during training, which impair their interpretability and generalization capability. To this end, we introduce ADAM, An emboDied causal Agent in Minecraft, which can autonomously navigate the open world, perceive multimodal context, learn causal world knowledge, and tackle complex tasks through lifelong learning. ADAM is empowered by four key components: 1) an interaction module, enabling the agent to execute actions while recording the interaction processes; 2) a causal model module, tasked with constructing an ever-growing causal graph from scratch, which enhances interpretability and reduces reliance on prior knowledge; 3) a controller module, comprising a planner, an actor, and a memory pool, using the learned causal graph to accomplish tasks; 4) a perception module, powered by multimodal large language models, enabling ADAM to perceive like a human player. Extensive experiments show that ADAM constructs a nearly perfect causal graph from scratch, enabling efficient task decomposition and execution with strong interpretability. Notably, in the modified Minecraft game where no prior knowledge is available, ADAM excels with remarkable robustness and generalization capability. ADAM pioneers a novel paradigm that integrates causal methods and embodied agents synergistically. Our project page is at https://opencausalab.github.io/ADAM.

3806. Euler Characteristic Tools for Topological Data Analysis

链接：https://iclr.cc/virtual/2025/poster/31379 abstract： In this article, we study Euler characteristic techniques in topological data analysis. Pointwise computing the Euler characteristic of a family of simplicial complexes built from data gives rise to the so-called Euler characteristic profile. We show that this simple descriptor achieves state-of-the-art performance in supervised tasks at a meagre computational cost. Inspired by signal analysis, we compute hybrid transforms of Euler characteristic profiles. These integral transforms mix Euler characteristic techniques with Lebesgue integration to provide highly efficient compressors of topological signals. As a consequence, they show remarkable performances in unsupervised settings. On the qualitative side, we provide numerous heuristics on the topological and geometric information captured by Euler profiles and their hybrid transforms. Finally, we prove stability results for these descriptors as well as asymptotic guarantees in random settings.

3807. Neural networks on Symmetric Spaces of Noncompact Type

链接：https://iclr.cc/virtual/2025/poster/29080 abstract： Recent works have demonstrated promising performances of neural networks on hyperbolic spaces and symmetric positive definite (SPD) manifolds. These spaces belong to a family of Riemannian manifolds referred to as symmetric spaces of noncompact type. In this paper, we propose a novel approach for developing neural networks on such spaces. Our approach relies on a unified formulation of the distance from a point to a hyperplane on the considered spaces. We show that some existing formulations of the point-to-hyperplane distance can be recovered by our approach under specific settings. Furthermore, we derive a closed-form expression for the point-to-hyperplane distance in higher-rank symmetric spaces of noncompact type equipped with G-invariant Riemannian metrics. The derived distance then serves as a tool to design fully-connected (FC) layers and an attention mechanism for neural networks on the considered spaces. Our approach is validated on challenging benchmarks for image classification, electroencephalogram (EEG) signal classification, image generation, and natural language inference.

3808. Exact Community Recovery under Side Information: Optimality of Spectral Algorithms

链接：https://iclr.cc/virtual/2025/poster/27656 abstract： We study the problem of exact community recovery in general, two-community block models, in the presence of node-attributed *side information*. We allow for a very general side information channel for node attributes, and for pairwise (edge) observations, consider both Bernoulli and Gaussian matrix models, capturing the Stochastic Block Model, Submatrix Localization, and $\mathbb{Z}_2$-Synchronization as special cases. A recent work of Dreveton et al. 2024 characterized the information-theoretic limit of a very general exact recovery problem with side information. In this paper, we show algorithmic achievability in the above important cases by designing a simple but optimal spectral algorithm that incorporates side information (when present) along with the eigenvectors of the pairwise observation matrix. Using the powerful tool of entrywise eigenvector analysis [Abbe et al. 2020], we show that our spectral algorithm can mimic the so called *genie-aided estimators*, where the i^{th} genie-aided estimator optimally computes the estimate of the i^{th} label, when all remaining labels are revealed by a genie. This perspective provides a unified understanding of the optimality of spectral algorithms for various exact recovery problems in a recent line of work.

3809. Content-Style Learning from Unaligned Domains: Identifiability under Unknown Latent Dimensions

链接：https://iclr.cc/virtual/2025/poster/28320 abstract： Understanding identifiability of latent content and style variables from unaligned multi-domain data is essential for tasks such asdomain translation and data generation. Existing works on content-style identification were often developed under somewhat stringent conditions, e.g., that all latent components are mutually independent and that the dimensions of the content and style variables are known. We introduce a new analytical framework via cross-domain latent distribution matching (LDM), which establishes content-style identifiability under substantially more relaxed conditions. Specifically, we show that restrictive assumptions such as component-wise independence of the latent variables can be removed. Most notably, we prove that prior knowledge of the content and style dimensions is not necessary for ensuring identifiability, if sparsity constraints are properly imposed onto the learned latent representations. Bypassing the knowledge of the exact latent dimension has been a longstanding aspiration in unsupervised representation learning---our analysis is the first to underpin its theoretical and practical viability. On the implementation side, we recast the LDM formulation into a regularized multi-domain GAN loss with coupled latent variables. We show that the reformulation is equivalent to LDM under mild conditions---yet requiring considerably less computational resource. Experiments corroborate with our theoretical claims.

3810. Let SSMs be ConvNets: State-space Modeling with Optimal Tensor Contractions

链接：https://iclr.cc/virtual/2025/poster/29733 abstract： We introduce Centaurus, a class of networks composed of generalized state-space model (SSM) blocks, where the SSM operations can be treated as tensor contractions during training. The optimal order of tensor contractions can then be systematically determined for every SSM block to maximize training efficiency. This allows more flexibility in designing SSM blocks beyond the depthwise-separable configuration commonly implemented. The new design choices will take inspiration from classical convolutional blocks including group convolutions, full convolutions, and bottleneck blocks. We architect the Centaurus network with a mixture of these blocks, to balance between network size and performance, as well as memory and computational efficiency during both training and inference. We show that this heterogeneous network design outperforms its homogeneous counterparts in raw audio processing tasks including keyword spotting, speech denoising, and automatic speech recognition (ASR). For ASR, Centaurus is the first network with competitive performance that can be made fully state-space based, without using any nonlinear recurrence (LSTMs), explicit convolutions (CNNs), or (surrogate) attention mechanism.

3811. Learning from End User Data with Shuffled Differential Privacy over Kernel Densities

链接：https://iclr.cc/virtual/2025/poster/29677 abstract： We study a setting of collecting and learning from private data distributed across end users.In the shuffled model of differential privacy, the end users partially protect their data locally before sharing it, and their data is also anonymized during its collection to enhance privacy. This model has recently become a prominent alternative to central DP, which requires full trust in a central data curator, and local DP, where fully local data protection takes a steep toll on downstream accuracy. Our main technical result is a shuffled DP protocol for privately estimating the kernel density function of a distributed dataset, with accuracy essentially matching central DP. We use it to privately learn a classifier from the end user data, by learning a private density function per class. Moreover, we show that the density function itself can recover the semantic content of its class, despite having been learned in the absence of any unprotected data. Our experiments show the favorable downstream performance of our approach, and highlight key downstream considerations and trade-offs in a practical ML deployment of shuffled DP.

3812. How to Verify Any (Reasonable) Distribution Property: Computationally Sound Argument Systems for Distributions

链接：https://iclr.cc/virtual/2025/poster/30266 abstract： As statistical analyses become more central to science, industry and society, there is a growing need to ensure correctness of their results. Approximate correctness can be verified by replicating the entire analysis, but can we verify without replication? We focus on distribution testing problems: verifying that an unknown distribution is close to having a claimed property. Our main contribution is an interactive protocol between a verifier and an untrusted prover, which can be used to verify any distribution property that can be decided in polynomial time given a full and explicit description of the distribution. If the distribution is at statistical distance ε from having the property, then the verifier rejects with high probability. This soundness property holds against any polynomial-time strategy that a cheating prover might follow, assuming the existence of collision-resistant hash functions (a standard assumption in cryptography). For distributions over a domain of size N, the protocol consists of 4 messages and the communication complexity and verifier runtime are roughly $\widetilde{O}\left(\sqrt{N} / \varepsilon^2 \right)$. The verifier's sample complexity is $\widetilde{O}\left(\sqrt{N} / \varepsilon^2 \right)$, and this is optimal up to $\text{polylog}(N)$ factors (for any protocol, regardless of its communication complexity). Even for simple properties, approximately deciding whether an unknown distribution has the property can require quasi-linear sample complexity and running time. For any such property, our protocol provides a quadratic speedup over replicating the analysis.

3813. ConceptPrune: Concept Editing in Diffusion Models via Skilled Neuron Pruning

链接：https://iclr.cc/virtual/2025/poster/28581 abstract： While large-scale text-to-image diffusion models have demonstrated impressive image-generation capabilities, there are significant concerns about their potential misuse for generating unsafe content, violating copyright, and perpetuating societal biases. Recently, the text-to-image generation community has begun addressing these concerns by editing or unlearning undesired concepts from pre-trained models. However, these methods often involve data-intensive and inefficient fine-tuning or utilize various forms of token remapping, rendering them susceptible to adversarial jailbreaks. In this paper, we present a simple and effective training-free approach, ConceptPrune, wherein we first identify critical regions within pre-trained models responsible for generating undesirable concepts, thereby facilitating straightforward concept unlearning via weight pruning. Experiments across a range of concepts including artistic styles, nudity, and object erasure demonstrate that target concepts can be efficiently erased by pruning a tiny fraction, approximately 0.12% of total weights, enabling multi-concept erasure and robustness against various white-box and black-box adversarial attacks.

3814. Deep MMD Gradient Flow without adversarial training

链接：https://iclr.cc/virtual/2025/poster/29740 abstract： We propose a gradient flow procedure for generative modeling by transporting particles from an initial source distribution to a target distribution, where the gradient field on the particles is given by a noise-adaptive Wasserstein Gradient of the Maximum Mean Discrepancy (MMD). The noise adaptive MMD is trained on data distributions corrupted by increasing levels of noise, obtained via a forward diffusion process, as commonly used in denoising diffusion probabilistic models. The result is a generalization of MMD Gradient Flow, which we call Diffusion-MMD-Gradient Flow or DMMD. The divergence training procedure is related to discriminator training in Generative Adversarial Networks (GAN), but does not require adversarial training. We obtain competitive empirical performance in unconditional image generation on CIFAR10, MNIST, CELEB-A (64 x64) and LSUN Church (64 x 64). Furthermore, we demonstrate the validity of the approach when MMD is replaced by a lower bound on the KL divergence.

3815. How Feature Learning Can Improve Neural Scaling Laws

链接：https://iclr.cc/virtual/2025/poster/28996 abstract： We develop a simple solvable model of neural scaling laws beyond the kernel limit. Theoretical analysis of this model predicts the performance scaling predictions with model size, training time and total amount of available data. From the scaling analysis we identify three relevant regimes: hard tasks, easy tasks, and super easy tasks. For easy and super-easy target functions, which are in the Hilbert space (RKHS) of the initial infinite-width neural tangent kernel (NTK), there is no change in the scaling exponents between feature learning models and models in the kernel regime. For hard tasks, which we define as tasks outside of the RKHS of the initial NTK, we show analytically and empirically that feature learning can improve the scaling with training time and compute, approximately doubling the exponent for very hard tasks. This leads to a new compute optimal scaling law for hard tasks in the feature learning regime. We support our finding that feature learning improves the scaling law for hard tasks with experiments of nonlinear MLPs fitting functions with power-law Fourier spectra on the circle and CNNs learning vision tasks.

3816. Understanding Factual Recall in Transformers via Associative Memories

链接：https://iclr.cc/virtual/2025/poster/28731 abstract： Large language models have demonstrated an impressive ability to perform factual recall. Prior work has found that transformers trained on factual recall tasks can store information at a rate proportional to their parameter count. In our work, we show that shallow transformers can use a combination of associative memories to obtain such near optimal storage capacity. We begin by proving that the storage capacities of both linear and MLP associative memories scale linearly with parameter count. We next introduce a synthetic factual recall task, and prove that a transformer with a single layer of self-attention followed by an MLP can obtain 100\% accuracy on the task whenever either the total number of self-attention parameters or MLP parameters scales (up to log factors) linearly with the number of facts. In particular, the transformer can trade off between using the value matrices or the MLP as an associative memory to store the dataset of facts. We complement these expressivity results with an analysis of the gradient flow trajectory of a simplified linear attention model trained on our factual recall task, where we show that the model exhibits sequential learning behavior.

3817. ViBiDSampler: Enhancing Video Interpolation Using Bidirectional Diffusion Sampler

链接：https://iclr.cc/virtual/2025/poster/28410 abstract： Recent progress in large-scale text-to-video (T2V) and image-to-video (I2V) diffusion models has greatly enhanced video generation, especially in terms of keyframe interpolation. However, current image-to-video diffusion models, while powerful in generating videos from a single conditioning frame, need adaptation for two-frame (start \& end) conditioned generation, which is essential for effective bounded interpolation. Unfortunately, existing approaches that fuse temporally forward and backward paths in parallel often suffer from off-manifold issues, leading to artifacts or requiring multiple iterative re-noising steps. In this work, we introduce a novel, bidirectional sampling strategy to address these off-manifold issues without requiring extensive re-noising or fine-tuning. Our method employs sequential sampling along both forward and backward paths, conditioned on the start and end frames, respectively, ensuring more coherent and on-manifold generation of intermediate frames. Additionally, we incorporate advanced guidance techniques, CFG++ and DDS, to further enhance the interpolation process. By integrating these, our method achieves state-of-the-art performance, efficiently generating high-quality, smooth videos between keyframes. On a single 3090 GPU, our method can interpolate 25 frames at 1024$\times$576 resolution in just 195 seconds, establishing it as a leading solution for keyframe interpolation.Project page: https://vibidsampler.github.io/

3818. Multi-LLM-Agents Debate - Performance, Efficiency, and Scaling Challenges

链接：https://iclr.cc/virtual/2025/poster/31346 abstract： Multi-Agent Debate (MAD) explores leveraging collaboration among multiple large language model (LLM) agents to improve test-time performance without additional training. This blog evaluates five MAD frameworks across nine benchmarks, revealing that current MAD methods fail to consistently outperform simpler single-agent strategies, even with increased computational resources. Analysis of factors such as agent configurations and debate rounds suggests that existing MAD designs fall short in fully utilizing additional inference-time computation.

3819. RTDiff: Reverse Trajectory Synthesis via Diffusion for Offline Reinforcement Learning

链接：https://iclr.cc/virtual/2025/poster/31264 abstract： In offline reinforcement learning (RL), managing the distribution shift between the learned policy and the static offline dataset is a persistent challenge that can result in overestimated values and suboptimal policies. Traditional offline RL methods address this by introducing conservative biases that limit exploration to well-understood regions, but they often overly restrict the agent's generalization capabilities. Recent work has sought to generate trajectories using generative models to augment the offline dataset, yet these methods still struggle with overestimating synthesized data, especially when out-of-distribution samples are produced. To overcome this issue, we propose RTDiff, a novel diffusion-based data augmentation technique that synthesizes trajectories in reverse, moving from unknown to known states. Such reverse generation naturally mitigates the risk of overestimation by ensuring that the agent avoids planning through unknown states. Additionally, reverse trajectory synthesis allows us to generate longer, more informative trajectories that take full advantage of diffusion models' generative strengths while ensuring reliability. We further enhance RTDiff by introducing flexible trajectory length control and improving the efficiency of the generation process through noise management. Our empirical results show that RTDiff significantly improves the performance of several state-of-the-art offline RL algorithms across diverse environments, achieving consistent and superior results by effectively overcoming distribution shift.

3820. Small-to-Large Generalization: Training Data Influences Models Consistently Across Scale

链接：https://iclr.cc/virtual/2025/poster/30845 abstract： Choice of training data distribution greatly influences model behavior. Yet, inlarge-scale settings, precisely characterizing how changes in trainingdata affects predictions is often difficult due to model training costs. Currentpractice is to instead extrapolate from scaled down, inexpensive-to-train proxymodels. However, changes in data do not influence smaller and larger modelsidentically. Therefore, understanding how choice of data affects large-scalemodels raises the question: how does training data distribution influence modelbehavior across compute scale? We find that small- and large-scale languagemodel predictions (generally) do highly correlate across choice oftraining data. Equipped with these findings, we characterize how proxy scaleaffects effectiveness in two downstream proxy model applications: dataattribution and dataset selection.

3821. Exploring The Forgetting in Adversarial Training: A Novel Method for Enhancing Robustness

链接：https://iclr.cc/virtual/2025/poster/28861 abstract： In recent years, there has been an explosion of research into developing robust deep neural networks against adversarial examples. As one of the most successful methods, Adversarial Training (AT) has been widely studied before, but there is still a gap to achieve promisingclean and robust accuracy for many practical tasks. In this paper, we consider the AT problem from a new perspective which connects it to catastrophic forgetting in continual learning (CL). Catastrophic forgetting is a phenomenon in which neural networks forget old knowledge upon learning a new task. Although AT and CL are two different problems, we show that they actually share several key properties in their training processes. Specifically, we conduct an empirical study and find that this forgetting phenomenon indeed occurs in adversarial robust training across multiple datasets (SVHN, CIFAR-10, CIFAR-100, and TinyImageNet) and perturbation models (ℓ_{∞} and ℓ_{2}). Based on this observation, we propose a novel method called Adaptive Multi-teachers Self-distillation (AMS), which leverages a carefully designed adaptive regularizer to mitigate the forgetting by aligning model outputs between new and old ``stages''. Moreover, our approach can be used as a unified method to enhance multiple different AT algorithms. Our experiments demonstrate that our method can significantly enhance robust accuracy and meanwhile preserve high clean accuracy, under several popular adversarial attacks (e.g., PGD, CW, and Auto Attacks). As another benefit of our method, we discover that it can largely alleviate the robust overfitting issue of AT in our experiments.

3822. Do LLM Agents Have Regret? A Case Study in Online Learning and Games

链接：https://iclr.cc/virtual/2025/poster/28236 abstract： Large language models (LLMs) have been increasingly employed for (interactive) decision-making, via the development of LLM-based autonomous agents. Despite their emerging successes, the performance of LLM agents in decision-making has not been fully investigated through quantitative metrics, especially in the multi-agent setting when they interact with each other, a typical scenario in real-world LLM-agent applications. To better understand the limits of LLM agents in these interactive environments, we propose to study their interactions in benchmark decision-making settings in online learning and game theory, through the performance metric of regret. We first empirically study the no-regret behaviors of LLMs in canonical non-stochastic online learning problems, as well as the emergence of equilibria when LLM agents interact through playing repeated games. We then provide some theoretical insights into the no-regret behaviors of LLM agents, under certain assumptions on the supervised pre-training and the rationality model of human decision-makers who generate the data. Notably, we also identify (simple) cases where advanced LLMs such as GPT-4 fail to be no-regret. To further promote the no-regret behaviors, we propose a novel unsupervised training loss of regret-loss, which, in contrast to the supervised pre-training loss, does not require the labels of (optimal) actions. Finally, we establish the statistical guarantee of generalization bound for regret-loss minimization, and more importantly, the optimization guarantee that minimizing such a loss may automatically lead to known no-regret learning algorithms, when single-layer self-attention models are used. Our further experiments demonstrate the effectiveness of our regret-loss, especially in addressing the above “regrettable” cases.

3823. Can LLM Simulations Truly Reflect Humanity? A Deep Dive

链接：https://iclr.cc/virtual/2025/poster/31340 abstract： Simulation powered by Large Language Models (LLMs) has become a promising method for exploring complex human social behaviors. However, the application of LLMs in simulations presents significant challenges, particularly regarding their capacity to accurately replicate the complexities of human behaviors and societal dynamics, as evidenced by recent studies highlighting discrepancies between simulated and real-world interactions. This blog rethinks LLM-based simulations by emphasizing both their limitations and the necessities for advancing LLM simulations. By critically examining these challenges, we aim to offer actionable insights and strategies for enhancing the applicability of LLM simulations in human society in the future.

3824. Compute-Constrained Data Selection

链接：https://iclr.cc/virtual/2025/poster/31001 abstract： Data selection can reduce the amount of training data needed to finetune LLMs; however, the efficacy of data selection scales directly with its compute. Motivated by the practical challenge of compute-constrained finetuning, we consider the setting in which both the cost of selecting data and training are budgeted for. We first formalize the problem of data selection with a cost-aware utility function, and model the data selection problem as trading off initial-selection cost for training gain. We run a comprehensive sweep of experiments across multiple tasks, varying compute budget by scaling finetuning tokens, model sizes, and data selection compute. Interestingly we find that many powerful data selection methods are almost never compute-optimal, and that cheaper data selection alternatives dominate both from a theoretical and empirical perspective. For compute-optimal training, we find that perplexity and gradient data selection require training-to-selection model size ratios of 5x and 10x, respectively.

3825. Multi-objective antibody design with constrained preference optimization

链接：https://iclr.cc/virtual/2025/poster/30996 abstract： Antibody design is crucial for developing therapies against diseases such as cancer and viral infections. Recent deep generative models have significantly advanced computational antibody design, particularly in enhancing binding affinity to target antigens. However, beyond binding affinity, antibodies should exhibit other favorable biophysical properties such as non-antigen binding specificity and low self-association, which are important for antibody developability and clinical safety. To address this challenge, we propose AbNovo, a framework that leverages constrained preference optimization for multi-objective antibody design. First, we pre-train an antigen-conditioned generative model for antibody structure and sequence co-design. Then, we fine-tune the model using binding affinity as a reward while enforcing explicit constraints on other biophysical properties. Specifically, we model the physical binding energy with continuous rewards rather than pairwise preferences and explore a primal-and-dual approach for constrained optimization. Additionally, we incorporate a structure-aware protein language model to mitigate the issue of limited training data. Evaluated on independent test sets, AbNovo outperforms existing methods in metrics of binding affinity such as Rosetta binding energy and evolutionary plausibility, as well as in metrics for other biophysical properties like stability and specificity.

3826. Linear SCM Identification in the Presence of Confounders and Gaussian Noise

链接：https://iclr.cc/virtual/2025/poster/29094 abstract： Noisy linear structural causal models (SCMs) in the presence of confounding variables are known to be identifiable if all confounding and noise variables are non-Gaussian and unidentifiable if all are Gaussian. The identifiability when only some are Gaussian remains concealed. We show that, in the presence of Gaussian noise, a linear SCM is uniquely identifiable provided that \emph{(i)} the number of confounders is at most the number of the observed variables, \emph{(ii)} the confounders do not have a Gaussian component, and \emph{(iii)} the causal structure of the SCM is known. If the third condition is relaxed, the SCM becomes finitely identifiable; more specifically, it belongs to a set of at most $n!$ linear SCMS, where n is the number of observed variables. The confounders in all of these $n!$ SCMs share the same joint probability distribution function (PDF), which we obtain analytically. For the case where both the noise and confounders are Gaussian, we provide further insight into the existing counter-example-based unidentifiability result and demonstrate that every SCM with confounders can be represented as an SCM without confounders but with the same joint PDF.

3827. No Location Left Behind: Measuring and Improving the Fairness of Implicit Representations for Earth Data

链接：https://iclr.cc/virtual/2025/poster/28761 abstract： Implicit neural representations (INRs) exhibit growing promise in addressing Earth representation challenges, ranging from emissions monitoring to climate modeling. However, existing methods disproportionately prioritize global average performance, whereas practitioners require fine-grained insights to understand biases and variations in these models. To bridge this gap, we introduce FAIR-Earth: a first-of-its-kind dataset explicitly crafted to challenge and examine inequities in Earth representations. FAIR-Earth comprises various high-resolution Earth signals, and uniquely aggregates extensive metadata along stratifications like landmass size and population density to assess the fairness of models. Evaluating state-of-the-art INRs across the various modalities of FAIR-Earth, we uncover striking performance disparities. Certain subgroups, especially those associated with high-frequency signals (e.g., islands, coastlines), are consistently poorly modeled by existing methods. In response, we propose spherical wavelet encodings, building on previous spatial encoding research for INRs. Leveraging the multi-resolution analysis capabilities of wavelets, our encodings yield more consistent performance over various scales and locations, offering more accurate and robust representations of the biased subgroups. These open-source contributions represent a crucial step towards facilitating the equitable assessment and deployment of implicit Earth representations.