Thu 2024.02.15

P-Mamba: Marrying Perona Malik Diffusion with Mamba for Efficient Pediatric Echocardiographic Left Ventricular Segmentation

Zi Ye, Tianxiang Chen

In pediatric cardiology, the accurate and immediate assessment of cardiac function through echocardiography is important since it can determine whether urgent intervention is required in many emergencies. However, echocardiography is characterized by ambiguity and heavy background noise interference, bringing more difficulty to accurate segmentation. Present methods lack efficiency and are also prone to mistakenly segmenting some background noise areas as the left ventricular area due to noise disturbance. To relieve the two issues, we introduce P-Mamba for efficient pediatric echocardiographic left ventricular segmentation. Specifically, we turn to the recently proposed vision mamba layers in our vision mamba encoder branch to improve the computing and memory efficiency of our model while modeling global dependencies. In the other DWT-based PMD encoder branch, we devise DWT-based Perona-Malik Diffusion (PMD) Blocks that utilize PMD for noise suppression, while simultaneously preserving the local shape cues of the left ventricle. Leveraging the strengths of both the two encoder branches, P-Mamba achieves superior accuracy and efficiency to established models, such as vision transformers with quadratic and linear computational complexity. This innovative approach promises significant advancements in pediatric cardiac imaging and beyond.

link: http://arxiv.org/abs/2402.08506v1

A PAC-Bayesian Link Between Generalisation and Flat Minima

Maxime Haddouche, Paul Viallard, Umut Simsekli, Benjamin Guedj

Modern machine learning usually involves predictors in the overparametrised setting (number of trained parameters greater than dataset size), and their training yield not only good performances on training data, but also good generalisation capacity. This phenomenon challenges many theoretical results, and remains an open problem. To reach a better understanding, we provide novel generalisation bounds involving gradient terms. To do so, we combine the PAC-Bayes toolbox with Poincar\'e and Log-Sobolev inequalities, avoiding an explicit dependency on dimension of the predictor space. Our results highlight the positive influence of \emph{flat minima} (being minima with a neighbourhood nearly minimising the learning problem as well) on generalisation performances, involving directly the benefits of the optimisation phase.

link: http://arxiv.org/abs/2402.08508v1

From Shapes to Shapes: Inferring SHACL Shapes for Results of SPARQL CONSTRUCT Queries (Extended Version)

Philipp Seifer, Daniel Hernández, Ralf Lämmel, Steffen Staab

SPARQL CONSTRUCT queries allow for the specification of data processing pipelines that transform given input graphs into new output graphs. It is now common to constrain graphs through SHACL shapes allowing users to understand which data they can expect and which not. However, it becomes challenging to understand what graph data can be expected at the end of a data processing pipeline without knowing the particular input data: Shape constraints on the input graph may affect the output graph, but may no longer apply literally, and new shapes may be imposed by the query template. In this paper, we study the derivation of shape constraints that hold on all possible output graphs of a given SPARQL CONSTRUCT query. We assume that the SPARQL CONSTRUCT query is fixed, e.g., being part of a program, whereas the input graphs adhere to input shape constraints but may otherwise vary over time and, thus, are mostly unknown. We study a fragment of SPARQL CONSTRUCT queries (SCCQ) and a fragment of SHACL (Simple SHACL). We formally define the problem of deriving the most restrictive set of Simple SHACL shapes that constrain the results from evaluating a SCCQ over any input graph restricted by a given set of Simple SHACL shapes. We propose and implement an algorithm that statically analyses input

SHACL shapes and CONSTRUCT queries and prove its soundness and complexity.

link: http://arxiv.org/abs/2402.08509v1

Amplifying Exploration in Monte-Carlo Tree Search by Focusing on the Unknown

Cedric Derstroff, Jannis Brugger, Jannis Blüml, Mira Mezini, Stefan Kramer, Kristian Kersting

Monte-Carlo tree search (MCTS) is an effective anytime algorithm with a vast amount of applications. It strategically allocates computational resources to focus on promising segments of the search tree, making it a very attractive search algorithm in large search spaces. However, it often expends its limited resources on reevaluating previously explored regions when they remain the most promising path. Our proposed methodology, denoted as AmEx-MCTS, solves this problem by introducing a novel MCTS formulation. Central to AmEx-MCTS is the decoupling of value updates, visit count updates, and the selected path during the tree search, thereby enabling the exclusion of already explored subtrees or leaves. This segregation preserves the utility of visit counts for both exploration-exploitation balancing and quality metrics within MCTS. The resultant augmentation facilitates in a considerably broader search using identical computational resources, preserving the essential characteristics of MCTS. The expanded coverage not only yields more precise estimations but also proves instrumental in larger and more complex problems. Our empirical evaluation demonstrates the superior performance of AmEx-MCTS, surpassing classical MCTS and related approaches by a substantial margin.

link: http://arxiv.org/abs/2402.08511v1

Counterfactual Influence in Markov Decision Processes

Milad Kazemi, Jessica Lally, Ekaterina Tishchenko, Hana Chockler, Nicola Paoletti

Our work addresses a fundamental problem in the context of counterfactual inference for Markov Decision Processes (MDPs). Given an MDP path τ, this kind of inference allows us to derive counterfactual paths τ' describing what-if versions of τ obtained under different action sequences than those observed in τ. However, as the counterfactual states and actions deviate from the observed ones over time, the observation τ may no longer influence the counterfactual world, meaning that the analysis is no longer tailored to the individual observation, resulting in interventional outcomes rather than counterfactual ones. Even though this issue specifically affects the popular Gumbel-max structural causal model used for MDP counterfactuals, it has remained overlooked until now. In this work, we introduce a formal characterisation of influence based on comparing counterfactual and interventional distributions. We devise an algorithm to construct counterfactual models that automatically satisfy influence constraints. Leveraging such models, we derive counterfactual policies that are not just optimal for a given reward structure but also remain tailored to the observed path. Even though there is an unavoidable trade-off between policy optimality and strength of influence constraints, our experiments demonstrate that it is possible to derive (near-)optimal policies while remaining under the influence of the observation.

link: http://arxiv.org/abs/2402.08514v1

Fairness Auditing with Multi-Agent Collaboration

Martijn de Vos, Akash Dhasade, Jade Garcia Bourrée, Anne-Marie Kermarrec, Erwan Le Merrer, Benoit Rottembourg, Gilles Tredan

Existing work in fairness audits assumes that agents operate independently. In this paper, we consider the case of multiple agents auditing the same platform for different tasks. Agents have two levers: their collaboration strategy, with or without coordination beforehand, and their sampling method. We theoretically study their interplay when agents operate independently or collaborate. We prove that, surprisingly, coordination can sometimes be detrimental to audit accuracy, whereas uncoordinated collaboration generally yields good results. Experimentation on real-world datasets confirms this observation, as the audit accuracy of uncoordinated collaboration matches that of collaborative optimal sampling.

link: http://arxiv.org/abs/2402.08522v1

Concept-1K: A Novel Benchmark for Instance Incremental Learning

Junhao Zheng, Shengjie Qiu, Qianli Ma

Incremental learning (IL) is essential to realize the human-level intelligence in the neural network. However, existing IL scenarios and datasets are unqualified for assessing forgetting in PLMs, giving an illusion that PLMs do not suffer from catastrophic forgetting. To this end, we propose a challenging IL scenario called instance-incremental learning (IIL) and a novel dataset called Concept-1K, which supports an order of magnitude larger IL steps. Based on the experiments on Concept-1K, we reveal that billion-parameter PLMs still suffer from catastrophic forgetting, and the forgetting is affected by both model scale, pretraining, and buffer size. Furthermore, existing IL methods and a popular finetuning technique, LoRA, fail to achieve satisfactory performance. Our study provides a novel scenario for future studies to explore the catastrophic forgetting of PLMs and encourage more powerful techniques to be designed for alleviating the forgetting in PLMs. The data, code and scripts are publicly available at https://github.com/zzz47zzz/pretrained-lm-for-incremental-learning.

link: http://arxiv.org/abs/2402.08526v1

Approximately Piecewise E(3) Equivariant Point Networks

Matan Atzmon, Jiahui Huang, Francis Williams, Or Litany

Integrating a notion of symmetry into point cloud neural networks is a provably effective way to improve their generalization capability. Of particular interest are $E(3)$ equivariant point cloud networks where Euclidean transformations applied to the inputs are preserved in the outputs. Recent efforts aim to extend networks that are $E(3)$ equivariant, to accommodate inputs made of multiple parts, each of which exhibits local $E(3)$ symmetry. In practical settings, however, the partitioning into individually transforming regions is unknown a priori. Errors in the partition prediction would unavoidably map to errors in respecting the true input symmetry. Past works have proposed different ways to predict the partition, which may exhibit uncontrolled errors in their ability to maintain equivariance to the actual partition. To this end, we introduce APEN: a general framework for constructing approximate piecewise-$E(3)$ equivariant point networks. Our primary insight is that functions that are equivariant with respect to a finer partition will also maintain equivariance in relation to the true partition. Leveraging this observation, we propose a design where the equivariance approximation error at each layers can be bounded solely in terms of (i) uncertainty quantification of the partition prediction, and (ii) bounds on the probability of failing to suggest a proper subpartition of the ground truth one. We demonstrate the effectiveness of APEN using two data types exemplifying part-based symmetry: (i) real-world scans of room scenes containing multiple furniture-type objects; and, (ii) human motions, characterized by articulated parts exhibiting rigid movement. Our empirical results demonstrate the advantage of integrating piecewise $E(3)$ symmetry into network design, showing a distinct improvement in generalization compared to prior works for both classification and segmentation tasks.

link: http://arxiv.org/abs/2402.08529v1

A Distributional Analogue to the Successor Representation

Harley Wiltzer, Jesse Farebrother, Arthur Gretton, Yunhao Tang, André Barreto, Will Dabney, Marc G. Bellemare, Mark Rowland

This paper contributes a new approach for distributional reinforcement learning which elucidates a clean separation of transition structure and reward in the learning process. Analogous to how the successor representation (SR) describes the expected consequences of behaving according to a given policy, our distributional successor measure (SM) describes the distributional consequences of this behaviour. We formulate the distributional SM as a distribution over distributions and provide theory connecting it with distributional and model-based reinforcement learning. Moreover, we propose an algorithm that learns the distributional SM from data by minimizing a two-level maximum mean discrepancy. Key to our method are a number of algorithmic techniques that are independently valuable for learning generative models of state. As an illustration of the usefulness of the distributional SM, we show that it enables zero-shot risk-sensitive policy evaluation in a way that was not previously possible.

link: http://arxiv.org/abs/2402.08530v1

Intelligent Diagnosis of Alzheimer's Disease Based on Machine Learning

Mingyang Li, Hongyu Liu, Yixuan Li, Zejun Wang, Yuan Yuan, Honglin Dai

This study is based on the Alzheimer's Disease Neuroimaging Initiative (ADNI) dataset and aims to explore early detection and disease progression in Alzheimer's disease (AD). We employ innovative data preprocessing strategies, including the use of the random forest algorithm to fill missing data and the handling of outliers and invalid data, thereby fully mining and utilizing these limited data resources. Through Spearman correlation coefficient analysis, we identify some features strongly correlated with AD diagnosis. We build and test three machine learning models using these features: random forest, XGBoost, and support vector machine (SVM). Among them, the XGBoost model performs the best in terms of diagnostic performance, achieving an accuracy of 91%. Overall, this study successfully overcomes the challenge of missing data and provides valuable insights into early detection of Alzheimer's disease, demonstrating its unique research value and practical significance.

link: http://arxiv.org/abs/2402.08539v1

Generative VS non-Generative Models in Engineering Shape Optimization

Muhammad Usama, Zahid Masood, Shahroz Khan, Konstantinos Kostas, Panagiotis Kaklis

In this work, we perform a systematic comparison of the effectiveness and efficiency of generative and non-generative models in constructing design spaces for novel and efficient design exploration and shape optimization. We apply these models in the case of airfoil/hydrofoil design and conduct the comparison on the resulting design spaces. A conventional Generative Adversarial Network (GAN) and a state-of-the-art generative model, the Performance-Augmented Diverse Generative Adversarial Network (PaDGAN), are juxtaposed with a linear non-generative model based on the coupling of the Karhunen-Lo\`eve Expansion and a physics-informed Shape Signature Vector (SSV-KLE). The comparison demonstrates that, with an appropriate shape encoding and a physics-augmented design space, non-generative models have the potential to cost-effectively generate high-performing valid designs with enhanced coverage of the design space. In this work, both approaches are applied to two large foil profile datasets comprising real-world and artificial designs generated through either a profile-generating parametric model or deep-learning approach. These datasets are further enriched with integral properties of their members' shapes as well as physics-informed parameters. Our results illustrate that the design spaces constructed by the non-generative model outperform the generative model in terms of design validity, generating robust latent spaces with none or significantly fewer invalid designs when compared to generative models. We aspire that these findings will aid the engineering design community in making informed decisions when constructing designs spaces for shape optimization, as we have show that under certain conditions computationally inexpensive approaches can closely match or even outperform state-of-the art generative models.

link: http://arxiv.org/abs/2402.08540v1

Dueling Over Dessert, Mastering the Art of Repeated Cake Cutting

Simina Brânzei, MohammadTaghi Hajiaghayi, Reed Phillips, Suho Shin, Kun Wang

We consider the setting of repeated fair division between two players, denoted Alice and Bob, with private valuations over a cake. In each round, a new cake arrives, which is identical to the ones in previous rounds. Alice cuts the cake at a point of her choice, while Bob chooses the left piece or the right piece, leaving the remainder for Alice. We consider two versions: sequential, where Bob observes Alice's cut point before choosing left/right, and simultaneous, where he only observes her cut point after making his choice. The simultaneous version was first considered by Aumann and Maschler (1995). We observe that if Bob is almost myopic and chooses his favorite piece too often, then he can be systematically exploited by Alice through a strategy akin to a binary search. This strategy allows Alice to approximate Bob's preferences with increasing precision, thereby securing a disproportionate share of the resource over time. We analyze the limits of how much a player can exploit the other one and show that fair utility profiles are in fact achievable. Specifically, the players can enforce the equitable utility profile of $(1/2, 1/2)$ in the limit on every trajectory of play, by keeping the other player's utility to approximately $1/2$ on average while guaranteeing they themselves get at least approximately $1/2$ on average. We show this theorem using a connection with Blackwell approachability. Finally, we analyze a natural dynamic known as fictitious play, where players best respond to the empirical distribution of the other player. We show that fictitious play converges to the equitable utility profile of $(1/2, 1/2)$ at a rate of $O(1/\sqrt{T})$.

link: http://arxiv.org/abs/2402.08547v1

Confronting Reward Overoptimization for Diffusion Models: A Perspective of Inductive and Primacy Biases

Ziyi Zhang, Sen Zhang, Yibing Zhan, Yong Luo, Yonggang Wen, Dacheng Tao

Bridging the gap between diffusion models and human preferences is crucial for their integration into practical generative workflows. While optimizing downstream reward models has emerged as a promising alignment strategy, concerns arise regarding the risk of excessive optimization with learned reward models, which potentially compromises ground-truth performance. In this work, we confront the reward overoptimization problem in diffusion model alignment through the lenses of both inductive and primacy biases. We first identify the divergence of current methods from the temporal inductive bias inherent in the multi-step denoising process of diffusion models as a potential source of overoptimization. Then, we surprisingly discover that dormant neurons in our critic model act as a regularization against overoptimization, while active neurons reflect primacy bias in this setting. Motivated by these observations, we propose Temporal Diffusion Policy Optimization with critic active neuron Reset (TDPO-R), a policy gradient algorithm that exploits the temporal inductive bias of intermediate timesteps, along with a novel reset strategy that targets active neurons to counteract the primacy bias. Empirical results demonstrate the superior efficacy of our algorithms in mitigating reward overoptimization.

link: http://arxiv.org/abs/2402.08552v1

Higher Layers Need More LoRA Experts

Chongyang Gao, Kezhen Chen, Jinmeng Rao, Baochen Sun, Ruibo Liu, Daiyi Peng, Yawen Zhang, Xiaoyuan Guo, Jie Yang, VS Subrahmanian

Parameter-efficient tuning (PEFT) techniques like low-rank adaptation (LoRA) offer training efficiency on Large Language Models, but their impact on model performance remains limited. Recent efforts integrate LoRA and Mixture-of-Experts (MoE) to improve the performance of PEFT methods. Despite promising results, research on improving the efficiency of LoRA with MoE is still in its early stages. Recent studies have shown that experts in the MoE architecture have different strengths and also exhibit some redundancy. Does this statement also apply to parameter-efficient MoE? In this paper, we introduce a novel parameter-efficient MoE method, \textit{\textbf{M}oE-L\textbf{o}RA with \textbf{L}ayer-wise Expert \textbf{A}llocation (MoLA)} for Transformer-based models, where each model layer has the flexibility to employ a varying number of LoRA experts. We investigate several architectures with varying layer-wise expert configurations. Experiments on six well-known NLP and commonsense QA benchmarks demonstrate that MoLA achieves equal or superior performance compared to all baselines. We find that allocating more LoRA experts to higher layers further enhances the effectiveness of models with a certain number of experts in total. With much fewer parameters, this allocation strategy outperforms the setting with the same number of experts in every layer. This work can be widely used as a plug-and-play parameter-efficient tuning approach for various applications. The code is available at https://github.com/GCYZSL/MoLA.

link: http://arxiv.org/abs/2402.08562v1

Denoising Diffusion Restoration Tackles Forward and Inverse Problems for the Laplace Operator

Amartya Mukherjee, Melissa M. Stadt, Lena Podina, Mohammad Kohandel, Jun Liu

Diffusion models have emerged as a promising class of generative models that map noisy inputs to realistic images. More recently, they have been employed to generate solutions to partial differential equations (PDEs). However, they still struggle with inverse problems in the Laplacian operator, for instance, the Poisson equation, because the eigenvalues that are large in magnitude amplify the measurement noise. This paper presents a novel approach for the inverse and forward solution of PDEs through the use of denoising diffusion restoration models (DDRM). DDRMs were used in linear inverse problems to restore original clean signals by exploiting the singular value decomposition (SVD) of the linear operator. Equivalently, we present an approach to restore the solution and the parameters in the Poisson equation by exploiting the eigenvalues and the eigenfunctions of the Laplacian operator. Our results show that using denoising diffusion restoration significantly improves the estimation of the solution and parameters. Our research, as a result, pioneers the integration of diffusion models with the principles of underlying physics to solve PDEs.

link: http://arxiv.org/abs/2402.08563v1