Sat 2024.03.16

Outlier Robust Multivariate Polynomial Regression

Vipul Arora, Arnab Bhattacharyya, Mathews Boban, Venkatesan Guruswami, Esty Kelman

We study the problem of robust multivariate polynomial regression: let $p\colon\mathbb{R}^n\to\mathbb{R}$ be an unknown n-variate polynomial of degree at most d in each variable. We are given as input a set of random samples $(\mathbf{x}_i,y_i) \in [-1,1]^n \times \mathbb{R}$ that are noisy versions of $(\mathbf{x}_i,p(\mathbf{x}_i))$. More precisely, each $\mathbf{x}_i$ is sampled independently from some distribution χ on $[-1,1]^n$, and for each i independently, y_i is arbitrary (i.e., an outlier) with probability at most $\rho < 1/2$, and otherwise satisfies $|y_i-p(\mathbf{x}_i)|\leq\sigma$. The goal is to output a polynomial $\hat{p}$, of degree at most d in each variable, within an ℓ_∞-distance of at most $O(\sigma)$ from p. Kane, Karmalkar, and Price [FOCS'17] solved this problem for $n=1$. We generalize their results to the n-variate setting, showing an algorithm that achieves a sample complexity of $O_n(d^n\log d)$, where the hidden constant depends on n, if χ is the n-dimensional Chebyshev distribution. The sample complexity is $O_n(d^{2n}\log d)$, if the samples are drawn from the uniform distribution instead. The approximation error is guaranteed to be at most $O(\sigma)$, and the run-time depends on $\log(1/\sigma)$. In the setting where each $\mathbf{x}_i$ and y_i are known up to N bits of precision, the run-time's dependence on N is linear. We also show that our sample complexities are optimal in terms of d^n. Furthermore, we show that it is possible to have the run-time be independent of $1/\sigma$, at the cost of a higher sample complexity.

link: http://arxiv.org/abs/2403.09465v1

Eta Inversion: Designing an Optimal Eta Function for Diffusion-based Real Image Editing

Wonjun Kang, Kevin Galim, Hyung Il Koo

Diffusion models have achieved remarkable success in the domain of text-guided image generation and, more recently, in text-guided image editing. A commonly adopted strategy for editing real images involves inverting the diffusion process to obtain a noisy representation of the original image, which is then denoised to achieve the desired edits. However, current methods for diffusion inversion often struggle to produce edits that are both faithful to the specified text prompt and closely resemble the source image. To overcome these limitations, we introduce a novel and adaptable diffusion inversion technique for real image editing, which is grounded in a theoretical analysis of the role of η in the DDIM sampling equation for enhanced editability. By designing a universal diffusion inversion method with a time- and region-dependent η function, we enable flexible control over the editing extent. Through a comprehensive series of quantitative and qualitative assessments, involving a comparison with a broad array of recent methods, we demonstrate the superiority of our approach. Our method not only sets a new benchmark in the field but also significantly outperforms existing strategies. Our code is available at https://github.com/furiosa-ai/eta-inversion

link: http://arxiv.org/abs/2403.09468v1

MambaTalk: Efficient Holistic Gesture Synthesis with Selective State Space Models

Zunnan Xu, Yukang Lin, Haonan Han, Sicheng Yang, Ronghui Li, Yachao Zhang, Xiu Li

Gesture synthesis is a vital realm of human-computer interaction, with wide-ranging applications across various fields like film, robotics, and virtual reality. Recent advancements have utilized the diffusion model and attention mechanisms to improve gesture synthesis. However, due to the high computational complexity of these techniques, generating long and diverse sequences with low latency remains a challenge. We explore the potential of state space models (SSMs) to address the challenge, implementing a two-stage modeling strategy with discrete motion priors to enhance the quality of gestures. Leveraging the foundational Mamba block, we introduce MambaTalk,

enhancing gesture diversity and rhythm through multimodal integration. Extensive experiments demonstrate that our method matches or exceeds the performance of state-of-the-art models.

link: http://arxiv.org/abs/2403.09471v1

Easy-to-Hard Generalization: Scalable Alignment Beyond Human Supervision

Zhiqing Sun, Longhui Yu, Yikang Shen, Weiyang Liu, Yiming Yang, Sean Welleck, Chuang Gan

Current AI alignment methodologies rely on human-provided demonstrations or judgments, and the learned capabilities of AI systems would be upper-bounded by human capabilities as a result. This raises a challenging research question: How can we keep improving the systems when their capabilities have surpassed the levels of humans? This paper answers this question in the context of tackling hard reasoning tasks (e.g., level 4-5 MATH problems) via learning from human annotations on easier tasks (e.g., level 1-3 MATH problems), which we term as \textit{easy-to-hard generalization}. Our key insight is that an evaluator (reward model) trained on supervisions for easier tasks can be effectively used for scoring candidate solutions of harder tasks and hence facilitating easy-to-hard generalization over different levels of tasks. Based on this insight, we propose a novel approach to scalable alignment, which firstly trains the process-supervised reward models on easy problems (e.g., level 1-3), and then uses them to evaluate the performance of policy models on hard problems. We show that such \textit{easy-to-hard generalization from evaluators} can enable \textit{easy-to-hard generalizations in generators} either through re-ranking or reinforcement learning (RL). Notably, our process-supervised 7b RL model achieves an accuracy of 34.0\% on MATH500, despite only using human supervision on easy problems. Our approach suggests a promising path toward AI systems that advance beyond the frontier of human supervision.

link: http://arxiv.org/abs/2403.09472v1

VIRUS-NeRF -- Vision, InfraRed and UltraSonic based Neural Radiance Fields

Nicolaj Schmid, Cornelius von Einem, Cesar Cadena, Roland Siegwart, Lorenz Hruby, Florian Tschopp

Autonomous mobile robots are an increasingly integral part of modern factory and warehouse operations. Obstacle detection, avoidance and path planning are critical safety-relevant tasks, which are often solved using expensive LiDAR sensors and depth cameras. We propose to use cost-effective low-resolution ranging sensors, such as ultrasonic and infrared time-of-flight sensors by developing VIRUS-NeRF - Vision, InfraRed, and UltraSonic based Neural Radiance Fields. Building upon Instant Neural Graphics Primitives with a Multiresolution Hash Encoding (Instant-NGP), VIRUS-NeRF incorporates depth measurements from ultrasonic and infrared sensors and utilizes them to update the occupancy grid used for ray marching. Experimental evaluation in 2D demonstrates that VIRUS-NeRF achieves comparable mapping performance to LiDAR point clouds regarding coverage. Notably, in small environments, its accuracy aligns with that of LiDAR measurements, while in larger ones, it is bounded by the utilized ultrasonic sensors. An in-depth ablation study reveals that adding ultrasonic and infrared sensors is highly effective when dealing with sparse data and low view variation. Further, the proposed occupancy grid of VIRUS-NeRF improves the mapping capabilities and increases the training speed by 46% compared to Instant-NGP. Overall, VIRUS-NeRF presents a promising approach for cost-effective local mapping in mobile robotics, with potential applications in safety and navigation tasks. The code can be found at https://github.com/ethz-asl/virus nerf.

link: http://arxiv.org/abs/2403.09477v1

Laying the Foundation First? Investigating the Generalization from Atomic Skills to Complex Reasoning Tasks

Yuncheng Huang, Qianyu He, Yipei Xu, Jiaqing Liang, Yanghua Xiao

Current language models have demonstrated their capability to develop basic reasoning, but struggle in more complicated reasoning tasks that require a combination of atomic skills, such as math word problem requiring skills like arithmetic and unit conversion. Previous methods either do

not improve the inherent atomic skills of models or not attempt to generalize the atomic skills to complex reasoning tasks. In this paper, we first propose a probing framework to investigate whether the atomic skill can spontaneously generalize to complex reasoning tasks. Then, we introduce a hierarchical curriculum learning training strategy to achieve better skill generalization. In our experiments, we find that atomic skills can not spontaneously generalize to compositional tasks. By leveraging hierarchical curriculum learning, we successfully induce generalization, significantly improve the performance of open-source LMs on complex reasoning tasks. Promisingly, the skill generalization exhibit effective in cross-dataset and cross-domain scenarios. Complex reasoning can also help enhance atomic skills. Our findings offer valuable guidance for designing better training strategies for complex reasoning tasks.

link: http://arxiv.org/abs/2403.09479v1

What Sketch Explainability Really Means for Downstream Tasks

Hmrishav Bandyopadhyay, Pinaki Nath Chowdhury, Ayan Kumar Bhunia, Aneeshan Sain, Tao Xiang, Yi-Zhe Song

In this paper, we explore the unique modality of sketch for explainability, emphasising the profound impact of human strokes compared to conventional pixel-oriented studies. Beyond explanations of network behavior, we discern the genuine implications of explainability across diverse downstream sketch-related tasks. We propose a lightweight and portable explainability solution -- a seamless plugin that integrates effortlessly with any pre-trained model, eliminating the need for re-training. Demonstrating its adaptability, we present four applications: highly studied retrieval and generation, and completely novel assisted drawing and sketch adversarial attacks. The centrepiece to our solution is a stroke-level attribution map that takes different forms when linked with downstream tasks. By addressing the inherent non-differentiability of rasterisation, we enable explanations at both coarse stroke level (SLA) and partial stroke level (P-SLA), each with its advantages for specific downstream tasks.

link: http://arxiv.org/abs/2403.09480v1

Clinical Reasoning over Tabular Data and Text with Bayesian Networks

Paloma Rabaey, Johannes Deleu, Stefan Heytens, Thomas Demeester

Bayesian networks are well-suited for clinical reasoning on tabular data, but are less compatible with natural language data, for which neural networks provide a successful framework. This paper compares and discusses strategies to augment Bayesian networks with neural text representations, both in a generative and discriminative manner. This is illustrated with simulation results for a primary care use case (diagnosis of pneumonia) and discussed in a broader clinical context.

link: http://arxiv.org/abs/2403.09481v1

SpikeReveal: Unlocking Temporal Sequences from Real Blurry Inputs with Spike Streams

Kang Chen, Shiyan Chen, Jiyuan Zhang, Baoyue Zhang, Yajing Zheng, Tiejun Huang, Zhaofei Yu

Reconstructing a sequence of sharp images from the blurry input is crucial for enhancing our insights into the captured scene and poses a significant challenge due to the limited temporal features embedded in the image. Spike cameras, sampling at rates up to 40,000 Hz, have proven effective in capturing motion features and beneficial for solving this ill-posed problem. Nonetheless, existing methods fall into the supervised learning paradigm, which suffers from notable performance degradation when applied to real-world scenarios that diverge from the synthetic training data domain. Moreover, the quality of reconstructed images is capped by the generated images based on motion analysis interpolation, which inherently differs from the actual scene, affecting the generalization ability of these methods in real high-speed scenarios. To address these challenges, we propose the first self-supervised framework for the task of spike-guided motion deblurring. Our approach begins with the formulation of a spike-guided deblurring model that explores the theoretical relationships among spike streams, blurry images, and their corresponding sharp sequences. We subsequently develop a self-supervised cascaded framework to alleviate the

issues of spike noise and spatial-resolution mismatching encountered in the deblurring model. With knowledge distillation and re-blurring loss, we further design a lightweight deblur network to generate high-quality sequences with brightness and texture consistency with the original input. Quantitative and qualitative experiments conducted on our real-world and synthetic datasets with spikes validate the superior generalization of the proposed framework. Our code, data and trained models will be available at \url{https://github.com/chenkang455/S-SDM}.

link: http://arxiv.org/abs/2403.09486v1

Rectifying Demonstration Shortcut in In-Context Learning

Joonwon Jang, Sanghwan Jang, Wonbin Kweon, Minjin Jeon, Hwanjo Yu

Large language models (LLMs) are able to solve various tasks with only a few demonstrations utilizing their in-context learning (ICL) abilities. However, LLMs often rely on their pre-trained semantic priors of demonstrations rather than on the input-label relationships to proceed with ICL prediction. In this work, we term this phenomenon as the `Demonstration Shortcut'. While previous works have primarily focused on improving ICL prediction results for predefined tasks, we aim to rectify the Demonstration Shortcut, thereby enabling the LLM to effectively learn new input-label relationships from demonstrations. To achieve this, we introduce In-Context Calibration, a demonstration-aware calibration method. We evaluate the effectiveness of the proposed method in two settings: (1) the Original ICL Task using the standard label space and (2) the Task Learning setting, where the label space is replaced with semantically unrelated tokens. In both settings, In-Context Calibration demonstrates substantial improvements, with results generalized across three LLM families (OPT, GPT, and Llama2) under various configurations.

link: http://arxiv.org/abs/2403.09488v1

Hyper-CL: Conditioning Sentence Representations with Hypernetworks

Young Hyun Yoo, Jii Cha, Changhyeon Kim, Taeuk Kim

While the introduction of contrastive learning frameworks in sentence representation learning has significantly contributed to advancements in the field, it still remains unclear whether state-of-the-art sentence embeddings can capture the fine-grained semantics of sentences, particularly when conditioned on specific perspectives. In this paper, we introduce Hyper-CL, an efficient methodology that integrates hypernetworks with contrastive learning to compute conditioned sentence representations. In our proposed approach, the hypernetwork is responsible for transforming pre-computed condition embeddings into corresponding projection layers. This enables the same sentence embeddings to be projected differently according to various conditions. Evaluation on two representative conditioning benchmarks, namely conditional semantic text similarity and knowledge graph completion, demonstrates that Hyper-CL is effective in flexibly conditioning sentence representations, showcasing its computational efficiency at the same time. We also provide a comprehensive analysis of the inner workings of our approach, leading to a better interpretation of its mechanisms.

link: http://arxiv.org/abs/2403.09490v1

On using Machine Learning Algorithms for Motorcycle Collision Detection

Philipp Rodegast, Steffen Maier, Jonas Kneifl, Jörg Fehr

Globally, motorcycles attract vast and varied users. However, since the rate of severe injury and fatality in motorcycle accidents far exceeds passenger car accidents, efforts have been directed toward increasing passive safety systems. Impact simulations show that the risk of severe injury or death in the event of a motorcycle-to-car impact can be greatly reduced if the motorcycle is equipped with passive safety measures such as airbags and seat belts. For the passive safety systems to be activated, a collision must be detected within milliseconds for a wide variety of impact configurations, but under no circumstances may it be falsely triggered. For the challenge of reliably detecting impending collisions, this paper presents an investigation towards the applicability of machine learning algorithms. First, a series of simulations of accidents and driving operation is introduced to collect data to train machine learning classification models. Their performance is

henceforth assessed and compared via multiple representative and application-oriented criteria.

link: http://arxiv.org/abs/2403.09491v1

Anomaly Detection by Adapting a pre-trained Vision Language Model

Yuxuan Cai, Xinwei He, Dingkang Liang, Ao Tong, Xiang Bai

Recently, large vision and language models have shown their success when adapting them to many downstream tasks. In this paper, we present a unified framework named CLIP-ADA for Anomaly Detection by Adapting a pre-trained CLIP model. To this end, we make two important improvements: 1) To acquire unified anomaly detection across industrial images of multiple categories, we introduce the learnable prompt and propose to associate it with abnormal patterns through self-supervised learning. 2) To fully exploit the representation power of CLIP, we introduce an anomaly region refinement strategy to refine the localization quality. During testing, the anomalies are localized by directly calculating the similarity between the representation of the learnable prompt and the image. Comprehensive experiments demonstrate the superiority of our framework, e.g., we achieve the state-of-the-art 97.5/55.6 and 89.3/33.1 on MVTec-AD and VisA for anomaly detection and localization. In addition, the proposed method also achieves encouraging performance with marginal training data, which is more challenging.

link: http://arxiv.org/abs/2403.09493v1

From Skepticism to Acceptance: Simulating the Attitude Dynamics Toward Fake News

Yuhan Liu, Xiuying Chen, Xiaoqing Zhang, Xing Gao, Ji Zhang, Rui Yan

In the digital era, the rapid propagation of fake news and rumors via social networks brings notable societal challenges and impacts public opinion regulation. Traditional fake news modeling typically forecasts the general popularity trends of different groups or numerically represents opinions shift. However, these methods often oversimplify real-world complexities and overlook the rich semantic information of news text. The advent of large language models (LLMs) provides the possibility of modeling subtle dynamics of opinion. Consequently, in this work, we introduce a Fake news Propagation Simulation framework (FPS) based on LLM, which studies the trends and control of fake news propagation in detail. Specifically, each agent in the simulation represents an individual with a distinct personality. They are equipped with both short-term and long-term memory, as well as a reflective mechanism to mimic human-like thinking. Every day, they engage in random opinion exchanges, reflect on their thinking, and update their opinions. Our simulation results uncover patterns in fake news propagation related to topic relevance, and individual traits, aligning with real-world observations. Additionally, we evaluate various intervention strategies and demonstrate that early and appropriately frequent interventions strike a balance between governance cost and effectiveness, offering valuable insights for practical applications. Our study underscores the significant utility and potential of LLMs in combating fake news.

link: http://arxiv.org/abs/2403.09498v1

A Reinforcement Learning Approach to Dairy Farm Battery Management using Q Learning

Nawazish Ali, Abdul Wahid, Rachael Shaw, Karl Mason

Dairy farming consumes a significant amount of energy, making it an energy-intensive sector within agriculture. Integrating renewable energy generation into dairy farming could help address this challenge. Effective battery management is important for integrating renewable energy generation. Managing battery charging and discharging poses significant challenges because of fluctuations in electrical consumption, the intermittent nature of renewable energy generation, and fluctuations in energy prices. Artificial Intelligence (AI) has the potential to significantly improve the use of renewable energy in dairy farming, however, there is limited research conducted in this particular domain. This research considers Ireland as a case study as it works towards attaining its 2030 energy strategy centered on the utilization of renewable sources. This study proposes a Q-learning-based algorithm for scheduling battery charging and discharging in a dairy farm setting.

This research also explores the effect of the proposed algorithm by adding wind generation data and considering additional case studies. The proposed algorithm reduces the cost of imported electricity from the grid by 13.41\%, peak demand by 2\%, and 24.49\% when utilizing wind generation. These results underline how reinforcement learning is highly effective in managing batteries in the dairy farming sector.

link: http://arxiv.org/abs/2403.09499v1