Thu 2024.03.14

On the Last-Iterate Convergence of Shuffling Gradient Methods

Zijian Liu, Zhengyuan Zhou

Shuffling gradient methods, which are also known as stochastic gradient descent (SGD) without replacement, are widely implemented in practice, particularly including three popular algorithms: Random Reshuffle (RR), Shuffle Once (SO), and Incremental Gradient (IG). Compared to the empirical success, the theoretical guarantee of shuffling gradient methods was not well-understanding for a long time. Until recently, the convergence rates had just been established for the average iterate for convex functions and the last iterate for strongly convex problems (using squared distance as the metric). However, when using the function value gap as the convergence criterion, existing theories cannot interpret the good performance of the last iterate in different settings (e.g., constrained optimization). To bridge this gap between practice and theory, we prove last-iterate convergence rates for shuffling gradient methods with respect to the objective value even without strong convexity. Our new results either (nearly) match the existing last-iterate lower bounds or are as fast as the previous best upper bounds for the average iterate.

link: http://arxiv.org/abs/2403.07723v1

Balancing Fairness and Accuracy in Data-Restricted Binary Classification

Zachary McBride Lazri, Danial Dervovic, Antigoni Polychroniadou, Ivan Brugere, Dana Dachman-Soled, Min Wu

Applications that deal with sensitive information may have restrictions placed on the data available to a machine learning (ML) classifier. For example, in some applications, a classifier may not have direct access to sensitive attributes, affecting its ability to produce accurate and fair decisions. This paper proposes a framework that models the trade-off between accuracy and fairness under four practical scenarios that dictate the type of data available for analysis. Prior works examine this trade-off by analyzing the outputs of a scoring function that has been trained to implicitly learn the underlying distribution of the feature vector, class label, and sensitive attribute of a dataset. In contrast, our framework directly analyzes the behavior of the optimal Bayesian classifier on this underlying distribution by constructing a discrete approximation it from the dataset itself. This approach enables us to formulate multiple convex optimization problems, which allow us to answer the question: How is the accuracy of a Bayesian classifier affected in different data restricting scenarios when constrained to be fair? Analysis is performed on a set of fairness definitions that include group and individual fairness. Experiments on three datasets demonstrate the utility of the proposed framework as a tool for quantifying the trade-offs among different fairness notions and their distributional dependencies.

link: http://arxiv.org/abs/2403.07724v1

SemEval-2024 Shared Task 6: SHROOM, a Shared-task on Hallucinations and Related Observable Overgeneration Mistakes

Timothee Mickus, Elaine Zosa, Raúl Vázquez, Teemu Vahtola, Jörg Tiedemann, Vincent Segonne, Alessandro Raganato, Marianna Apidianaki

This paper presents the results of the SHROOM, a shared task focused on detecting hallucinations: outputs from natural language generation (NLG) systems that are fluent, yet inaccurate. Such cases of overgeneration put in jeopardy many NLG applications, where correctness is often mission-critical. The shared task was conducted with a newly constructed dataset of 4000 model outputs labeled by 5 annotators each, spanning 3 NLP tasks: machine translation, paraphrase generation and definition modeling. The shared task was tackled by a total of 58 different users grouped in 42 teams, out of which 27 elected to write a system description paper; collectively, they submitted over 300 prediction sets on both tracks of the shared task. We observe a number of key trends in how this approach was tackled -- many participants rely on a handful of model, and often rely either on synthetic data for fine-tuning or zero-shot prompting strategies. While a majority of the

teams did outperform our proposed baseline system, the performances of top-scoring systems are still consistent with a random handling of the more challenging items.

link: http://arxiv.org/abs/2403.07726v1

CAS: A General Algorithm for Online Selective Conformal Prediction with FCR Control

Yajie Bao, Yuyang Huo, Haojie Ren, Changliang Zou

We study the problem of post-selection predictive inference in an online fashion. To avoid devoting resources to unimportant units, a preliminary selection of the current individual before reporting its prediction interval is common and meaningful in online predictive tasks. Since the online selection causes a temporal multiplicity in the selected prediction intervals, it is important to control the real-time false coverage-statement rate (FCR) to measure the averaged miscoverage error. We develop a general framework named CAS (Calibration after Adaptive Selection) that can wrap around any prediction model and online selection rule to output post-selection prediction intervals. If the current individual is selected, we first perform an adaptive selection on historical data to construct a calibration set, then output a conformal prediction interval for the unobserved label. We provide tractable constructions for the calibration set for popular online selection rules. We proved that CAS can achieve an exact selection-conditional coverage guarantee in the finite-sample and distribution-free regimes. For the decision-driven selection rule, including most online multiple-testing procedures, CAS can exactly control the real-time FCR below the target level without any distributional assumptions. For the online selection with symmetric thresholds, we establish the error bound for the control gap of FCR under mild distributional assumptions. To account for the distribution shift in online data, we also embed CAS into some recent dynamic conformal prediction methods and examine the long-run FCR control. Numerical results on both synthetic and real data corroborate that CAS can effectively control FCR around the target level and yield more narrowed prediction intervals over existing baselines across various settings.

link: http://arxiv.org/abs/2403.07728v1

DSEG-LIME - Improving Image Explanation by Hierarchical Data-Driven Segmentation

Patrick Knab, Sascha Marton, Christian Bartelt

Explainable Artificial Intelligence is critical in unraveling decision-making processes in complex machine learning models. LIME (Local Interpretable Model-agnostic Explanations) is a well-known XAI framework for image analysis. It utilizes image segmentation to create features to identify relevant areas for classification. Consequently, poor segmentation can compromise the consistency of the explanation and undermine the importance of the segments, affecting the overall interpretability. Addressing these challenges, we introduce DSEG-LIME (Data-Driven Segmentation LIME), featuring: i) a data-driven segmentation for human-recognized feature generation, and ii) a hierarchical segmentation procedure through composition. We benchmark DSEG-LIME on pre-trained models with images from the ImageNet dataset - scenarios without domain-specific knowledge. The analysis includes a quantitative evaluation using established XAI metrics, complemented by a qualitative assessment through a user study. Our findings demonstrate that DSEG outperforms in most of the XAI metrics and enhances the alignment of explanations with human-recognized concepts, significantly improving interpretability. The code is available under: https://github. com/patrick-knab/DSEG-LIME

link: http://arxiv.org/abs/2403.07733v1

The Minimax Rate of HSIC Estimation for Translation-Invariant Kernels

Florian Kalinke, Zoltan Szabo

Kernel techniques are among the most influential approaches in data science and statistics. Under mild conditions, the reproducing kernel Hilbert space associated to a kernel is capable of encoding the independence of $M\ge 2$ random variables. Probably the most widespread independence measure relying on kernels is the so-called Hilbert-Schmidt independence criterion (HSIC; also

referred to as distance covariance in the statistics literature). Despite various existing HSIC estimators designed since its introduction close to two decades ago, the fundamental question of the rate at which HSIC can be estimated is still open. In this work, we prove that the minimax optimal rate of HSIC estimation on $\mathbb R^d$ for Borel measures containing the Gaussians with continuous bounded translation-invariant characteristic kernels is $\mathcal O\!\left(n^{-1/2}\right)$. Specifically, our result implies the optimality in the minimax sense of many of the most-frequently used estimators (including the U-statistic, the V-statistic, and the Nystr\"om-based one) on $\mathbb R^d$.

link: http://arxiv.org/abs/2403.07735v1

Uncertainty Quantification with Deep Ensembles for 6D Object Pose Estimation

Kira Wursthorn, Markus Hillemann, Markus Ulrich

The estimation of 6D object poses is a fundamental task in many computer vision applications. Particularly, in high risk scenarios such as human-robot interaction, industrial inspection, and automation, reliable pose estimates are crucial. In the last years, increasingly accurate and robust deep-learning-based approaches for 6D object pose estimation have been proposed. Many top-performing methods are not end-to-end trainable but consist of multiple stages. In the context of deep uncertainty quantification, deep ensembles are considered as state of the art since they have been proven to produce well-calibrated and robust uncertainty estimates. However, deep ensembles can only be applied to methods that can be trained end-to-end. In this work, we propose a method to quantify the uncertainty of multi-stage 6D object pose estimation approaches with deep ensembles. For the implementation, we choose SurfEmb as representative, since it is one of the top-performing 6D object pose estimation approaches in the BOP Challenge 2022. We apply established metrics and concepts for deep uncertainty quantification to evaluate the results. Furthermore, we propose a novel uncertainty calibration score for regression tasks to quantify the quality of the estimated uncertainty.

link: http://arxiv.org/abs/2403.07741v1

Equipping Computational Pathology Systems with Artifact Processing Pipelines: A Showcase for Computation and Performance Trade-offs

Neel Kanwal, Farbod Khoraminia, Umay Kiraz, Andres Mosquera-Zamudio, Carlos Monteagudo, Emiel A. M. Janssen, Tahlita C. M. Zuiverloon, Chunmig Rong, Kjersti Engan

Histopathology is a gold standard for cancer diagnosis under a microscopic examination. However, histological tissue processing procedures result in artifacts, which are ultimately transferred to the digitized version of glass slides, known as whole slide images (WSIs). Artifacts are diagnostically irrelevant areas and may result in wrong deep learning (DL) algorithms predictions. Therefore, detecting and excluding artifacts in the computational pathology (CPATH) system is essential for reliable automated diagnosis. In this paper, we propose a mixture of experts (MoE) scheme for detecting five notable artifacts, including damaged tissue, blur, folded tissue, air bubbles, and histologically irrelevant blood from WSIs. First, we train independent binary DL models as experts to capture particular artifact morphology. Then, we ensemble their predictions using a fusion mechanism. We apply probabilistic thresholding over the final probability distribution to improve the sensitivity of the MoE. We developed DL pipelines using two MoEs and two multiclass models of state-of-the-art deep convolutional neural networks (DCNNs) and vision transformers (ViTs). DCNNs-based MoE and ViTs-based MoE schemes outperformed simpler multiclass models and were tested on datasets from different hospitals and cancer types, where MoE using DCNNs yielded the best results. The proposed MoE yields 86.15% F1 and 97.93% sensitivity scores on unseen data, retaining less computational cost for inference than MoE using ViTs. This best performance of MoEs comes with relatively higher computational trade-offs than multiclass models. The proposed artifact detection pipeline will not only ensure reliable CPATH predictions but may also provide quality control.

link: http://arxiv.org/abs/2403.07743v2

Probabilistic Easy Variational Causal Effect

Usef Faghihi, Amir Saki

Let X and Z be random vectors, and $Y=g(X,Z)$. In this paper, on the one hand, for the case that X and Z are continuous, by using the ideas from the total variation and the flux of g, we develop a point of view in causal inference capable of dealing with a broad domain of causal problems. Indeed, we focus on a function, called Probabilistic Easy Variational Causal Effect (PEACE), which can measure the direct causal effect of X on Y with respect to continuously and interventionally changing the values of X while keeping the value of Z constant. PEACE is a function of $d\ge 0$, which is a degree managing the strengths of probability density values $f(x|z)$. On the other hand, we generalize the above idea for the discrete case and show its compatibility with the continuous case. Further, we investigate some properties of PEACE using measure theoretical concepts. Furthermore, we provide some identifiability criteria and several examples showing the generic capability of PEACE. We note that PEACE can deal with the causal problems for which micro-level or just macro-level changes in the value of the input variables are important. Finally, PEACE is stable under small changes in $\partial g_{in}/\partial x$ and the joint distribution of X and Z, where g_{in} is obtained from g by removing all functional relationships defining X and Z.

link: http://arxiv.org/abs/2403.07745v1

Unleashing HyDRa: Hybrid Fusion, Depth Consistency and Radar for Unified 3D Perception

Philipp Wolters, Johannes Gilg, Torben Teepe, Fabian Herzog, Anouar Laouichi, Martin Hofmann, Gerhard Rigoll

Low-cost, vision-centric 3D perception systems for autonomous driving have made significant progress in recent years, narrowing the gap to expensive LiDAR-based methods. The primary challenge in becoming a fully reliable alternative lies in robust depth prediction capabilities, as camera-based systems struggle with long detection ranges and adverse lighting and weather conditions. In this work, we introduce HyDRa, a novel camera-radar fusion architecture for diverse 3D perception tasks. Building upon the principles of dense BEV (Bird's Eye View)-based architectures, HyDRa introduces a hybrid fusion approach to combine the strengths of complementary camera and radar features in two distinct representation spaces. Our Height Association Transformer module leverages radar features already in the perspective view to produce more robust and accurate depth predictions. In the BEV, we refine the initial sparse representation by a Radar-weighted Depth Consistency. HyDRa achieves a new state-of-the-art for camera-radar fusion of 64.2 NDS (+1.8) and 58.4 AMOTA (+1.5) on the public nuScenes dataset. Moreover, our new semantically rich and spatially accurate BEV features can be directly converted into a powerful occupancy representation, beating all previous camera-based methods on the Occ3D benchmark by an impressive 3.7 mIoU.

link: http://arxiv.org/abs/2403.07746v1

FineMath: A Fine-Grained Mathematical Evaluation Benchmark for Chinese Large Language Models

Yan Liu, Renren Jin, Lin Shi, Zheng Yao, Deyi Xiong

To thoroughly assess the mathematical reasoning abilities of Large Language Models (LLMs), we need to carefully curate evaluation datasets covering diverse mathematical concepts and mathematical problems at different difficulty levels. In pursuit of this objective, we propose FineMath in this paper, a fine-grained mathematical evaluation benchmark dataset for assessing Chinese LLMs. FineMath is created to cover the major key mathematical concepts taught in elementary school math, which are further divided into 17 categories of math word problems, enabling in-depth analysis of mathematical reasoning abilities of LLMs. All the 17 categories of math word problems are manually annotated with their difficulty levels according to the number of reasoning steps required to solve these problems. We conduct extensive experiments on a wide range of LLMs on FineMath and find that there is still considerable room for improvements in terms of mathematical reasoning capability of Chinese LLMs. We also carry out an in-depth analysis on the evaluation

process and methods that have been overlooked previously. These two factors significantly influence the model results and our understanding of their mathematical reasoning capabilities. The dataset will be publicly available soon.

link: http://arxiv.org/abs/2403.07747v1

Ariadne and Theseus: Exploration and Rendezvous with Two Mobile Agents in an Unknown Graph

Romain Cosson

We investigate two fundamental problems in mobile computing: exploration and rendezvous, with two distinct mobile agents in an unknown graph. The agents can read and write information on whiteboards that are located at all nodes. They both move along one adjacent edge at every time-step. In the exploration problem, both agents start from the same node of the graph and must traverse all of its edges. We show that a simple variant of depth-first search achieves collective exploration in m synchronous time-steps, where m is the number of edges of the graph. This improves the competitive ratio of collective graph exploration. In the rendezvous problem, the agents start from different nodes of the graph and must meet as fast as possible. We introduce an algorithm guaranteeing rendezvous in at most $\frac{3}{2}m$ time-steps. This improves over the so-called `wait for Mommy' algorithm which requires $2m$ time-steps. All our guarantees are derived from a more general asynchronous setting in which the speeds of the agents are controlled by an adversary at all times. Our guarantees also generalize to weighted graphs, if the number of edges m is replaced by the sum of all edge lengths.

link: http://arxiv.org/abs/2403.07748v1

Synth^2: Boosting Visual-Language Models with Synthetic Captions and Image Embeddings

Sahand Sharifzadeh, Christos Kaplanis, Shreya Pathak, Dharshan Kumaran, Anastasija Ilic, Jovana Mitrovic, Charles Blundell, Andrea Banino

The creation of high-quality human-labeled image-caption datasets presents a significant bottleneck in the development of Visual-Language Models (VLMs). We propose a novel approach that leverages the strengths of Large Language Models (LLMs) and image generation models to create synthetic image-text pairs for efficient and effective VLM training. Our method employs pretraining a text-to-image model to synthesize image embeddings starting from captions generated by an LLM. These synthetic pairs are then used to train a VLM. Extensive experiments demonstrate that the VLM trained with synthetic data exhibits comparable performance on image captioning, while requiring a fraction of the data used by models trained solely on human-annotated data. In particular, we outperform the baseline by 17% through augmentation with a synthetic dataset. Furthermore, we show that synthesizing in the image embedding space is 25% faster than in the pixel space. This research introduces a promising technique for generating large-scale, customizable image datasets, leading to enhanced VLM performance and wider applicability across various domains, all with improved data efficiency and resource utilization.

link: http://arxiv.org/abs/2403.07750v1

Vision-based Vehicle Re-identification in Bridge Scenario using Flock Similarity

Chunfeng Zhang, Ping Wang

Due to the needs of road traffic flow monitoring and public safety management, video surveillance cameras are widely distributed in urban roads. However, the information captured directly by each camera is siloed, making it difficult to use it effectively. Vehicle re-identification refers to finding a vehicle that appears under one camera in another camera, which can correlate the information captured by multiple cameras. While license plate recognition plays an important role in some applications, there are some scenarios where re-identification method based on vehicle appearance are more suitable. The main challenge is that the data of vehicle appearance has the characteristics of high inter-class similarity and large intra-class differences. Therefore, it is difficult to accurately distinguish between different vehicles by relying only on vehicle appearance

information. At this time, it is often necessary to introduce some extra information, such as spatio-temporal information. Nevertheless, the relative position of the vehicles rarely changes when passing through two adjacent cameras in the bridge scenario. In this paper, we present a vehicle re-identification method based on flock similarity, which improves the accuracy of vehicle re-identification by utilizing vehicle information adjacent to the target vehicle. When the relative position of the vehicles remains unchanged and flock size is appropriate, we obtain an average relative improvement of 204% on VeRi dataset in our experiments. Then, the effect of the magnitude of the relative position change of the vehicles as they pass through two cameras is discussed. We present two metrics that can be used to quantify the difference and establish a connection between them. Although this assumption is based on the bridge scenario, it is often true in other scenarios due to driving safety and camera location.

link: http://arxiv.org/abs/2403.07752v1

Emerging Technologies for 6G Non-Terrestrial-Networks: From Academia to Industrial Applications

Cong T. Nguyen, Yuris Mulya Saputra, Nguyen Van Huynh, Tan N. Nguyen, Dinh Thai Hoang, Diep N Nguyen, Van-Quan Pham, Miroslav Voznak, Symeon Chatzinotas, Dinh-Hieu Tran

Terrestrial networks form the fundamental infrastructure of modern communication systems, serving more than 4 billion users globally. However, terrestrial networks are facing a wide range of challenges, from coverage and reliability to interference and congestion. As the demands of the 6G era are expected to be much higher, it is crucial to address these challenges to ensure a robust and efficient communication infrastructure for the future. To address these problems, Non-terrestrial Network (NTN) has emerged to be a promising solution. NTNs are communication networks that leverage airborne (e.g., unmanned aerial vehicles) and spaceborne vehicles (e.g., satellites) to facilitate ultra-reliable communications and connectivity with high data rates and low latency over expansive regions. This article aims to provide a comprehensive survey on the utilization of network slicing, Artificial Intelligence/Machine Learning (AI/ML), and Open Radio Access Network (ORAN) to address diverse challenges of NTNs from the perspectives of both academia and industry. Particularly, we first provide an in-depth tutorial on NTN and the key enabling technologies including network slicing, AI/ML, and ORAN. Then, we provide a comprehensive survey on how network slicing and AI/ML have been leveraged to overcome the challenges that NTNs are facing. Moreover, we present how ORAN can be utilized for NTNs. Finally, we highlight important challenges, open issues, and future research directions of NTN in the 6G era.

link: http://arxiv.org/abs/2403.07763v1