Sat 2024.02.24

Overcoming Dimensional Collapse in Self-supervised Contrastive Learning for Medical Image Segmentation

Jamshid Hassanpour, Vinkle Srivastav, Didier Mutter, Nicolas Padoy

Self-supervised learning (SSL) approaches have achieved great success when the amount of labeled data is limited. Within SSL, models learn robust feature representations by solving pretext tasks. One such pretext task is contrastive learning, which involves forming pairs of similar and dissimilar input samples, guiding the model to distinguish between them. In this work, we investigate the application of contrastive learning to the domain of medical image analysis. Our findings reveal that MoCo v2, a state-of-the-art contrastive learning method, encounters dimensional collapse when applied to medical images. This is attributed to the high degree of inter-image similarity shared between the medical images. To address this, we propose two key contributions: local feature learning and feature decorrelation. Local feature learning improves the ability of the model to focus on the local regions of the image, while feature decorrelation removes the linear dependence among the features. Our experimental findings demonstrate that our contributions significantly enhance the model's performance in the downstream task of medical segmentation, both in the linear evaluation and full fine-tuning settings. This work illustrates the importance of effectively adapting SSL techniques to the characteristics of medical imaging tasks.

link: http://arxiv.org/abs/2402.14611v1

Two Counterexamples to \textit{Tokenization and the Noiseless Channel}

Marco Cognetta, Vilém Zouhar, Sangwhan Moon, Naoaki Okazaki

In \textit{Tokenization and the Noiseless Channel} \cite{zouhar-etal-2023-tokenization}, R\'enyi efficiency is suggested as an intrinsic mechanism for evaluating a tokenizer: for NLP tasks, the tokenizer which leads to the highest R\'enyi efficiency of the unigram distribution should be chosen. The R\'enyi efficiency is thus treated as a predictor of downstream performance (e.g., predicting BLEU for a machine translation task), without the expensive step of training multiple models with different tokenizers. Although useful, the predictive power of this metric is not perfect, and the authors note there are additional qualities of a good tokenization scheme that R\'enyi efficiency alone cannot capture. We describe two variants of BPE tokenization which can arbitrarily increase R\'enyi efficiency while decreasing the downstream model performance. These counterexamples expose cases where R\'enyi efficiency fails as an intrinsic tokenization metric and thus give insight for building more accurate predictors.

link: http://arxiv.org/abs/2402.14614v1

The Impact of Word Splitting on the Semantic Content of Contextualized Word Representations

Aina Garí Soler, Matthieu Labeau, Chloé Clavel

When deriving contextualized word representations from language models, a decision needs to be made on how to obtain one for out-of-vocabulary (OOV) words that are segmented into subwords. What is the best way to represent these words with a single vector, and are these representations of worse quality than those of in-vocabulary words? We carry out an intrinsic evaluation of embeddings from different models on semantic similarity tasks involving OOV words. Our analysis reveals, among other interesting findings, that the quality of representations of words that are split is often, but not always, worse than that of the embeddings of known words. Their similarity values, however, must be interpreted with caution.

link: http://arxiv.org/abs/2402.14616v1

Seer: Proactive Revenue-Aware Scheduling for Live Streaming Services in Crowdsourced Cloud-Edge Platforms

Shaoyuan Huang, Zheng Wang, Zhongtian Zhang, Heng Zhang, Xiaofei Wang, Wenyu Wang

As live streaming services skyrocket, Crowdsourced Cloud-edge service Platforms (CCPs) have surfaced as pivotal intermediaries catering to the mounting demand. Despite the role of stream scheduling to CCPs' Quality of Service (QoS) and throughput, conventional optimization strategies struggle to enhancing CCPs' revenue, primarily due to the intricate relationship between resource utilization and revenue. Additionally, the substantial scale of CCPs magnifies the difficulties of time-intensive scheduling. To tackle these challenges, we propose Seer, a proactive revenue-aware scheduling system for live streaming services in CCPs. The design of Seer is motivated by meticulous measurements of real-world CCPs environments, which allows us to achieve accurate revenue modeling and overcome three key obstacles that hinder the integration of prediction and optimal scheduling. Utilizing an innovative Pre-schedule-Execute-Re-schedule paradigm and flexible scheduling modes, Seer achieves efficient revenue-optimized scheduling in CCPs. Extensive evaluations demonstrate Seer's superiority over competitors in terms of revenue, utilization, and anomaly penalty mitigation, boosting CCPs revenue by 147% and expediting scheduling $3.4 \times$ faster.

link: http://arxiv.org/abs/2402.14619v1

latrend: A Framework for Clustering Longitudinal Data

Niek Den Teuling, Steffen Pauws, Edwin van den Heuvel

Clustering of longitudinal data is used to explore common trends among subjects over time for a numeric measurement of interest. Various R packages have been introduced throughout the years for identifying clusters of longitudinal patterns, summarizing the variability in trajectories between subject in terms of one or more trends. We introduce the R package "latrend" as a framework for the unified application of methods for longitudinal clustering, enabling comparisons between methods with minimal coding. The package also serves as an interface to commonly used packages for clustering longitudinal data, including "dtwclust", "flexmix", "kml", "lcmm", "mclust", "mixAK", and "mixtools". This enables researchers to easily compare different approaches, implementations, and method specifications. Furthermore, researchers can build upon the standard tools provided by the framework to quickly implement new cluster methods, enabling rapid prototyping. We demonstrate the functionality and application of the latrend package on a synthetic dataset based on the therapy adherence patterns of patients with sleep apnea.

link: http://arxiv.org/abs/2402.14621v1

From Keywords to Structured Summaries: Streamlining Scholarly Knowledge Access

Mahsa Shamsabadi, Jennifer D'Souza

This short paper highlights the growing importance of information retrieval (IR) engines in the scientific community, addressing the inefficiency of traditional keyword-based search engines due to the rising volume of publications. The proposed solution involves structured records, underpinning advanced information technology (IT) tools, including visualization dashboards, to revolutionize how researchers access and filter articles, replacing the traditional text-heavy approach. This vision is exemplified through a proof of concept centered on the ``reproductive number estimate of infectious diseases'' research theme, using a fine-tuned large language model (LLM) to automate the creation of structured records to populate a backend database that now goes beyond keywords. The result is a next-generation IR method accessible at https://orkg.org/usecases/r0-estimates.

link: http://arxiv.org/abs/2402.14622v1

RoboScript: Code Generation for Free-Form Manipulation Tasks across Real and Simulation

Junting Chen, Yao Mu, Qiaojun Yu, Tianming Wei, Silang Wu, Zhecheng Yuan, Zhixuan Liang, Chao Yang, Kaipeng Zhang, Wenqi Shao, Yu Qiao, Huazhe Xu, Mingyu Ding, Ping Luo

Rapid progress in high-level task planning and code generation for open-world robot manipulation has been witnessed in Embodied AI. However, previous studies put much effort into general common sense reasoning and task planning capabilities of large-scale language or multi-modal models, relatively little effort on ensuring the deployability of generated code on real robots, and other fundamental components of autonomous robot systems including robot perception, motion planning, and control. To bridge this ``ideal-to-real'' gap, this paper presents \textbf{RobotScript}, a platform for 1) a deployable robot manipulation pipeline powered by code generation; and 2) a code generation benchmark for robot manipulation tasks in free-form natural language. The RobotScript platform addresses this gap by emphasizing the unified interface with both simulation and real robots, based on abstraction from the Robot Operating System (ROS), ensuring syntax compliance and simulation validation with Gazebo. We demonstrate the adaptability of our code generation framework across multiple robot embodiments, including the Franka and UR5 robot arms, and multiple grippers. Additionally, our benchmark assesses reasoning abilities for physical space and constraints, highlighting the differences between GPT-3.5, GPT-4, and Gemini in handling complex physical interactions. Finally, we present a thorough evaluation on the whole system, exploring how each module in the pipeline: code generation, perception, motion planning, and even object geometric properties, impact the overall performance of the system.

link: http://arxiv.org/abs/2402.14623v1

Measuring NAT64 Usage in the Wild

Elizabeth Boswell, Stephen McQuistin, Colin Perkins, Stephen Strowes

NAT64 is an IPv6 transition mechanism that enables IPv6-only hosts to access the IPv4 Internet. Understanding the deployment of NAT64, and its performance impact, is crucial to the success of the IPv6 transition, by encouraging IPv6-only deployments. We develop a set of tests for detecting NAT64 and apply them to the RIPE Atlas testbed, finding 224 probes, in 43 networks, that can use NAT64 to access the IPv4 Internet. Using 34 dual stack probes, that have both NAT64 and native IPv4 access, to compare performance, we find that NAT64 paths are, on average, 23.13% longer, with 17.47% higher round-trip times.

link: http://arxiv.org/abs/2402.14632v1

Distributed Radiance Fields for Edge Video Compression and Metaverse Integration in Autonomous Driving

Eugen Šlapak, Matúš Dopiriak, Mohammad Abdullah Al Faruque, Juraj Gazda, Marco Levorato

The metaverse is a virtual space that combines physical and digital elements, creating immersive and connected digital worlds. For autonomous mobility, it enables new possibilities with edge computing and digital twins (DTs) that offer virtual prototyping, prediction, and more. DTs can be created with 3D scene reconstruction methods that capture the real world's geometry, appearance, and dynamics. However, sending data for real-time DT updates in the metaverse, such as camera images and videos from connected autonomous vehicles (CAVs) to edge servers, can increase network congestion, costs, and latency, affecting metaverse services. Herein, a new method is proposed based on distributed radiance fields (RFs), multi-access edge computing (MEC) network for video compression and metaverse DT updates. RF-based encoder and decoder are used to create and restore representations of camera images. The method is evaluated on a dataset of camera images from the CARLA simulator. Data savings of up to 80% were achieved for H.264 I-frame - P-frame pairs by using RFs instead of I-frames, while maintaining high peak signal-to-noise ratio (PSNR) and structural similarity index measure (SSIM) qualitative metrics for the reconstructed images. Possible uses and challenges for the metaverse and autonomous mobility are also discussed.

link: http://arxiv.org/abs/2402.14642v1

Sparse Linear Regression and Lattice Problems

Aparna Gupte, Neekon Vafa, Vinod Vaikuntanathan

Sparse linear regression (SLR) is a well-studied problem in statistics where one is given a design matrix $X\in\mathbb{R}^{m\times n}$ and a response vector $y=X\theta^*+w$ for a k-sparse vector θ^* (that is, $\|\theta^*\|_0\leq k$) and small, arbitrary noise w, and the goal is to find a k-sparse $\widehat{\theta} \in \mathbb{R}^n$ that minimizes the mean squared prediction error $\frac{1}{m}\|X\widehat{\theta}-X\theta^*\|^2_2$. While ℓ_1-relaxation methods such as basis pursuit, Lasso, and the Dantzig selector solve SLR when the design matrix is well-conditioned, no general algorithm is known, nor is there any formal evidence of hardness in an average-case setting with respect to all efficient algorithms. We give evidence of average-case hardness of SLR w.r.t. all efficient algorithms assuming the worst-case hardness of lattice problems. Specifically, we give an instance-by-instance reduction from a variant of the bounded distance decoding (BDD) problem on lattices to SLR, where the condition number of the lattice basis that defines the BDD instance is directly related to the restricted eigenvalue condition of the design matrix, which characterizes some of the classical statistical-computational gaps for sparse linear regression. Also, by appealing to worst-case to average-case reductions from the world of lattices, this shows hardness for a distribution of SLR instances; while the design matrices are ill-conditioned, the resulting SLR instances are in the identifiable regime. Furthermore, for well-conditioned (essentially) isotropic Gaussian design matrices, where Lasso is known to behave well in the identifiable regime, we show hardness of outputting any good solution in the unidentifiable regime where there are many solutions, assuming the worst-case hardness of standard and well-studied lattice problems.

link: http://arxiv.org/abs/2402.14645v1

CoLoRA: Continuous low-rank adaptation for reduced implicit neural modeling of parameterized partial differential equations

Jules Berman, Benjamin Peherstorfer

This work introduces reduced models based on Continuous Low Rank Adaptation (CoLoRA) that pre-train neural networks for a given partial differential equation and then continuously adapt low-rank weights in time to rapidly predict the evolution of solution fields at new physics parameters and new initial conditions. The adaptation can be either purely data-driven or via an equation-driven variational approach that provides Galerkin-optimal approximations. Because CoLoRA approximates solution fields locally in time, the rank of the weights can be kept small, which means that only few training trajectories are required offline so that CoLoRA is well suited for data-scarce regimes. Predictions with CoLoRA are orders of magnitude faster than with classical methods and their accuracy and parameter efficiency is higher compared to other neural network approaches.

link: http://arxiv.org/abs/2402.14646v1

Rethinking Invariance Regularization in Adversarial Training to Improve Robustness-Accuracy Trade-off

Futa Waseda, Isao Echizen

Although adversarial training has been the state-of-the-art approach to defend against adversarial examples (AEs), they suffer from a robustness-accuracy trade-off. In this work, we revisit representation-based invariance regularization to learn discriminative yet adversarially invariant representations, aiming to mitigate this trade-off. We empirically identify two key issues hindering invariance regularization: (1) a "gradient conflict" between invariance loss and classification objectives, indicating the existence of "collapsing solutions," and (2) the mixture distribution problem arising from diverged distributions of clean and adversarial inputs. To address these issues, we propose Asymmetrically Representation-regularized Adversarial Training (AR-AT), which incorporates a stop-gradient operation and a pre-dictor in the invariance loss to avoid "collapsing solutions," inspired by a recent non-contrastive self-supervised learning approach, and a split-BatchNorm (BN) structure to resolve the mixture distribution problem. Our method significantly improves the robustness-accuracy trade-off by learning adversarially invariant representations without sacrificing discriminative power. Furthermore, we discuss the relevance of our findings to knowledge-distillation-based defense methods, contributing to a deeper understanding of their relative successes.

link: http://arxiv.org/abs/2402.14648v1

GaussianPro: 3D Gaussian Splatting with Progressive Propagation

Kai Cheng, Xiaoxiao Long, Kaizhi Yang, Yao Yao, Wei Yin, Yuexin Ma, Wenping Wang, Xuejin Chen

The advent of 3D Gaussian Splatting (3DGS) has recently brought about a revolution in the field of neural rendering, facilitating high-quality renderings at real-time speed. However, 3DGS heavily depends on the initialized point cloud produced by Structure-from-Motion (SfM) techniques. When tackling with large-scale scenes that unavoidably contain texture-less surfaces, the SfM techniques always fail to produce enough points in these surfaces and cannot provide good initialization for 3DGS. As a result, 3DGS suffers from difficult optimization and low-quality renderings. In this paper, inspired by classical multi-view stereo (MVS) techniques, we propose GaussianPro, a novel method that applies a progressive propagation strategy to guide the densification of the 3D Gaussians. Compared to the simple split and clone strategies used in 3DGS, our method leverages the priors of the existing reconstructed geometries of the scene and patch matching techniques to produce new Gaussians with accurate positions and orientations. Experiments on both large-scale and small-scale scenes validate the effectiveness of our method, where our method significantly surpasses 3DGS on the Waymo dataset, exhibiting an improvement of 1.15dB in terms of PSNR.

link: http://arxiv.org/abs/2402.14650v1

Cleaner Pretraining Corpus Curation with Neural Web Scraping

Zhipeng Xu, Zhenghao Liu, Yukun Yan, Zhiyuan Liu, Chenyan Xiong, Ge Yu

The web contains large-scale, diverse, and abundant information to satisfy the information-seeking needs of humans. Through meticulous data collection, preprocessing, and curation, webpages can be used as a fundamental data resource for language model pretraining. However, when confronted with the progressively revolutionized and intricate nature of webpages, rule-based/feature-based web scrapers are becoming increasingly inadequate. This paper presents a simple, fast, and effective Neural web Scraper (NeuScraper) to help extract primary and clean text contents from webpages. Experimental results show that NeuScraper surpasses the baseline scrapers by achieving more than a 20% improvement, demonstrating its potential in extracting higher-quality data to facilitate the language model pretraining. All of the code is available at https://github.com/OpenMatch/NeuScraper.

link: http://arxiv.org/abs/2402.14652v1

Multi-HMR: Multi-Person Whole-Body Human Mesh Recovery in a Single Shot

Fabien Baradel, Matthieu Armando, Salma Galaaoui, Romain Brégier, Philippe Weinzaepfel, Grégory Rogez, Thomas Lucas

We present Multi-HMR, a strong single-shot model for multi-person 3D human mesh recovery from a single RGB image. Predictions encompass the whole body, i.e, including hands and facial expressions, using the SMPL-X parametric model and spatial location in the camera coordinate system. Our model detects people by predicting coarse 2D heatmaps of person centers, using features produced by a standard Vision Transformer (ViT) backbone. It then predicts their whole-body pose, shape and spatial location using a new cross-attention module called the Human Prediction Head (HPH), with one query per detected center token, attending to the entire set of features. As direct prediction of SMPL-X parameters yields suboptimal results, we introduce CUFFS; the Close-Up Frames of Full-Body Subjects dataset, containing humans close to the camera with diverse hand poses. We show that incorporating this dataset into training further enhances predictions, particularly for hands, enabling us to achieve state-of-the-art performance. Multi-HMR also optionally accounts for camera intrinsics, if available, by encoding camera ray directions for each image token. This simple design achieves strong performance on whole-body and body-only benchmarks simultaneously. We train models with various backbone sizes and input resolutions. In particular, using a ViT-S backbone and 448×448 input images already yields a fast and competitive model with respect to state-of-the-art methods, while considering larger

models and higher resolutions further improve performance.
link: http://arxiv.org/abs/2402.14654v1