Fri 2024.02.16

Spectral Filters, Dark Signals, and Attention Sinks

Nicola Cancedda

Projecting intermediate representations onto the vocabulary is an increasingly popular interpretation tool for transformer-based LLMs, also known as the logit lens. We propose a quantitative extension to this approach and define spectral filters on intermediate representations based on partitioning the singular vectors of the vocabulary embedding and unembedding matrices into bands. We find that the signals exchanged in the tail end of the spectrum are responsible for attention sinking (Xiao et al. 2023), of which we provide an explanation. We find that the loss of pretrained models can be kept low despite suppressing sizable parts of the embedding spectrum in a layer-dependent way, as long as attention sinking is preserved. Finally, we discover that the representation of tokens that draw attention from many tokens have large projections on the tail end of the spectrum.

link: http://arxiv.org/abs/2402.09221v1

Is my Data in your AI Model? Membership Inference Test with Application to Face Images

Daniel DeAlcala, Aythami Morales, Gonzalo Mancera, Julian Fierrez, Ruben Tolosana, Javier Ortega-Garcia

This paper introduces the Membership Inference Test (MINT), a novel approach that aims to empirically assess if specific data was used during the training of Artificial Intelligence (AI) models. Specifically, we propose two novel MINT architectures designed to learn the distinct activation patterns that emerge when an audited model is exposed to data used during its training process. The first architecture is based on a Multilayer Perceptron (MLP) network and the second one is based on Convolutional Neural Networks (CNNs). The proposed MINT architectures are evaluated on a challenging face recognition task, considering three state-of-the-art face recognition models. Experiments are carried out using six publicly available databases, comprising over 22 million face images in total. Also, different experimental scenarios are considered depending on the context available of the AI model to test. Promising results, up to 90% accuracy, are achieved using our proposed MINT approach, suggesting that it is possible to recognize if an AI model has been trained with specific data.

link: http://arxiv.org/abs/2402.09225v1

Directional Convergence Near Small Initializations and Saddles in Two-Homogeneous Neural Networks

Akshay Kumar, Jarvis Haupt

This paper examines gradient flow dynamics of two-homogeneous neural networks for small initializations, where all weights are initialized near the origin. For both square and logistic losses, it is shown that for sufficiently small initializations, the gradient flow dynamics spend sufficient time in the neighborhood of the origin to allow the weights of the neural network to approximately converge in direction to the Karush-Kuhn-Tucker (KKT) points of a neural correlation function that quantifies the correlation between the output of the neural network and corresponding labels in the training data set. For square loss, it has been observed that neural networks undergo saddle-to-saddle dynamics when initialized close to the origin. Motivated by this, this paper also shows a similar directional convergence among weights of small magnitude in the neighborhood of certain saddle points.

link: http://arxiv.org/abs/2402.09226v1

Context Composing for Full Line Code Completion

Anton Semenkin, Yaroslav Sokolov, Evgeniia Vu

Code Completion is one of the most used Integrated Development Environment (IDE) features, which affects the everyday life of a software developer. Modern code completion approaches moved from the composition of several static analysis-based contributors to pipelines that involve neural networks. This change allows the proposal of longer code suggestions while maintaining the relatively short time spent on generation itself. At JetBrains, we put a lot of effort into perfecting the code completion workflow so it can be both helpful and non-distracting for a programmer. We managed to ship the Full Line Code Completion feature to PyCharm Pro IDE and proved its usefulness in A/B testing on hundreds of real Python users. The paper describes our approach to context composing for the Transformer model that is a core of the feature's implementation. In addition to that, we share our next steps to improve the feature and emphasize the importance of several research aspects in the area.

link: http://dx.doi.org/10.1145/3643796.3648446

Investigating Premature Convergence in Co-optimization of Morphology and Control in Evolved Virtual Soft Robots

Alican Mertan, Nick Cheney

Evolving virtual creatures is a field with a rich history and recently it has been getting more attention, especially in the soft robotics domain. The compliance of soft materials endows soft robots with complex behavior, but it also makes their design process unintuitive and in need of automated design. Despite the great interest, evolved virtual soft robots lack the complexity, and co-optimization of morphology and control remains a challenging problem. Prior work identifies and investigates a major issue with the co-optimization process -- fragile co-adaptation of brain and body resulting in premature convergence of morphology. In this work, we expand the investigation of this phenomenon by comparing learnable controllers with proprioceptive observations and fixed controllers without any observations, whereas in the latter case, we only have the optimization of the morphology. Our experiments in two morphology spaces and two environments that vary in complexity show, concrete examples of the existence of high-performing regions in the morphology space that are not able to be discovered during the co-optimization of the morphology and control, yet exist and are easily findable when optimizing morphologies alone. Thus this work clearly demonstrates and characterizes the challenges of optimizing morphology during co-optimization. Based on these results, we propose a new body-centric framework to think about the co-optimization problem which helps us understand the issue from a search perspective. We hope the insights we share with this work attract more attention to the problem and help us to enable efficient brain-body co-optimization.

link: http://arxiv.org/abs/2402.09231v1

Design and Realization of a Benchmarking Testbed for Evaluating Autonomous Platooning Algorithms

Michael Shaham, Risha Ranjan, Engin Kirda, Taskin Padir

Autonomous vehicle platoons present near- and long-term opportunities to enhance operational efficiencies and save lives. The past 30 years have seen rapid development in the autonomous driving space, enabling new technologies that will alleviate the strain placed on human drivers and reduce vehicle emissions. This paper introduces a testbed for evaluating and benchmarking platooning algorithms on 1/10th scale vehicles with onboard sensors. To demonstrate the testbed's utility, we evaluate three algorithms, linear feedback and two variations of distributed model predictive control, and compare their results on a typical platooning scenario where the lead vehicle tracks a reference trajectory that changes speed multiple times. We validate our algorithms in simulation to analyze the performance as the platoon size increases, and find that the distributed model predictive control algorithms outperform linear feedback on hardware and in simulation.

link: http://arxiv.org/abs/2402.09233v1

Multi-Hierarchical Surrogate Learning for Structural Dynamical Crash Simulations Using Graph Convolutional Neural Networks

Jonas Kneifl, Jörg Fehr, Steven L. Brunton, J. Nathan Kutz

Crash simulations play an essential role in improving vehicle safety, design optimization, and injury risk estimation. Unfortunately, numerical solutions of such problems using state-of-the-art high-fidelity models require significant computational effort. Conventional data-driven surrogate modeling approaches create low-dimensional embeddings for evolving the dynamics in order to circumvent this computational effort. Most approaches directly operate on high-resolution data obtained from numerical discretization, which is both costly and complicated for mapping the flow of information over large spatial distances. Furthermore, working with a fixed resolution prevents the adaptation of surrogate models to environments with variable computing capacities, different visualization resolutions, and different accuracy requirements. We thus propose a multi-hierarchical framework for structurally creating a series of surrogate models for a kart frame, which is a good proxy for industrial-relevant crash simulations, at different levels of resolution. For multiscale phenomena, macroscale features are captured on a coarse surrogate, whereas microscale effects are resolved by finer ones. The learned behavior of the individual surrogates is passed from coarse to finer levels through transfer learning. In detail, we perform a mesh simplification on the kart model to obtain multi-resolution representations of it. We then train a graph-convolutional neural network-based surrogate that learns parameter-dependent low-dimensional latent dynamics on the coarsest representation. Subsequently, another, similarly structured surrogate is trained on the residual of the first surrogate using a finer resolution. This step can be repeated multiple times. By doing so, we construct multiple surrogates for the same system with varying hardware requirements and increasing accuracy.

link: http://arxiv.org/abs/2402.09234v2

Learning Interpretable Concepts: Unifying Causal Representation Learning and Foundation Models

Goutham Rajendran, Simon Buchholz, Bryon Aragam, Bernhard Schölkopf, Pradeep Ravikumar

To build intelligent machine learning systems, there are two broad approaches. One approach is to build inherently interpretable models, as endeavored by the growing field of causal representation learning. The other approach is to build highly-performant foundation models and then invest efforts into understanding how they work. In this work, we relate these two approaches and study how to learn human-interpretable concepts from data. Weaving together ideas from both fields, we formally define a notion of concepts and show that they can be provably recovered from diverse data. Experiments on synthetic data and large language models show the utility of our unified approach.

link: http://arxiv.org/abs/2402.09236v1

Weatherproofing Retrieval for Localization with Generative AI and Geometric Consistency

Yannis Kalantidis, Mert Bülent Sar■y■ld■z, Rafael S. Rezende, Philippe Weinzaepfel, Diane Larlus, Gabriela Csurka

State-of-the-art visual localization approaches generally rely on a first image retrieval step whose role is crucial. Yet, retrieval often struggles when facing varying conditions, due to e.g. weather or time of day, with dramatic consequences on the visual localization accuracy. In this paper, we improve this retrieval step and tailor it to the final localization task. Among the several changes we advocate for, we propose to synthesize variants of the training set images, obtained from generative text-to-image models, in order to automatically expand the training set towards a number of nameable variations that particularly hurt visual localization. After expanding the training set, we propose a training approach that leverages the specificities and the underlying geometry of this mix of real and synthetic images. We experimentally show that those changes translate into large improvements for the most challenging visual localization datasets. Project page: https://europe.naverlabs.com/ret4loc

link: http://arxiv.org/abs/2402.09237v1

Robust Training of Temporal GNNs using Nearest Neighbours based Hard Negatives

Shubham Gupta, Srikanta Bedathur

Temporal graph neural networks Tgnn have exhibited state-of-art performance in future-link prediction tasks. Training of these TGNNs is enumerated by uniform random sampling based unsupervised loss. During training, in the context of a positive example, the loss is computed over uninformative negatives, which introduces redundancy and sub-optimal performance. In this paper, we propose modified unsupervised learning of Tgnn, by replacing the uniform negative sampling with importance-based negative sampling. We theoretically motivate and define the dynamically computed distribution for a sampling of negative examples. Finally, using empirical evaluations over three real-world datasets, we show that Tgnn trained using loss based on proposed negative sampling provides consistent superior performance.

link: http://dx.doi.org/10.1145/3632410.3632464

Switch EMA: A Free Lunch for Better Flatness and Sharpness

Siyuan Li, Zicheng Liu, Juanxi Tian, Ge Wang, Zedong Wang, Weiyang Jin, Di Wu, Cheng Tan, Tao Lin, Yang Liu, Baigui Sun, Stan Z. Li

Exponential Moving Average (EMA) is a widely used weight averaging (WA) regularization to learn flat optima for better generalizations without extra cost in deep neural network (DNN) optimization. Despite achieving better flatness, existing WA methods might fall into worse final performances or require extra test-time computations. This work unveils the full potential of EMA with a single line of modification, i.e., switching the EMA parameters to the original model after each epoch, dubbed as Switch EMA (SEMA). From both theoretical and empirical aspects, we demonstrate that SEMA can help DNNs to reach generalization optima that better trade-off between flatness and sharpness. To verify the effectiveness of SEMA, we conduct comparison experiments with discriminative, generative, and regression tasks on vision and language datasets, including image classification, self-supervised learning, object detection and segmentation, image generation, video prediction, attribute regression, and language modeling. Comprehensive results with popular optimizers and networks show that SEMA is a free lunch for DNN training by improving performances and boosting convergence speeds.

link: http://arxiv.org/abs/2402.09240v1

Efficient One-stage Video Object Detection by Exploiting Temporal Consistency

Guanxiong Sun, Yang Hua, Guosheng Hu, Neil Robertson

Recently, one-stage detectors have achieved competitive accuracy and faster speed compared with traditional two-stage detectors on image data. However, in the field of video object detection (VOD), most existing VOD methods are still based on two-stage detectors. Moreover, directly adapting existing VOD methods to one-stage detectors introduces unaffordable computational costs. In this paper, we first analyse the computational bottlenecks of using one-stage detectors for VOD. Based on the analysis, we present a simple yet efficient framework to address the computational bottlenecks and achieve efficient one-stage VOD by exploiting the temporal consistency in video frames. Specifically, our method consists of a location-prior network to filter out background regions and a size-prior network to skip unnecessary computations on low-level feature maps for specific frames. We test our method on various modern one-stage detectors and conduct extensive experiments on the ImageNet VID dataset. Excellent experimental results demonstrate the superior effectiveness, efficiency, and compatibility of our method. The code is available at https://github.com/guanxiongsun/vfe.pytorch.

link: http://dx.doi.org/10.1007/978-3-031-19833-5_1

Synthesizing Knowledge-enhanced Features for Real-world Zero-shot Food Detection

Pengfei Zhou, Weiqing Min, Jiajun Song, Yang Zhang, Shuqiang Jiang

Food computing brings various perspectives to computer vision like vision-based food analysis for nutrition and health. As a fundamental task in food computing, food detection needs Zero-Shot Detection (ZSD) on novel unseen food objects to support real-world scenarios, such as intelligent kitchens and smart restaurants. Therefore, we first benchmark the task of Zero-Shot Food Detection (ZSFD) by introducing FOWA dataset with rich attribute annotations. Unlike ZSD, fine-grained problems in ZSFD like inter-class similarity make synthesized features inseparable. The complexity of food semantic attributes further makes it more difficult for current ZSD methods to distinguish various food categories. To address these problems, we propose a novel framework ZSFDet to tackle fine-grained problems by exploiting the interaction between complex attributes. Specifically, we model the correlation between food categories and attributes in ZSFDet by multi-source graphs to provide prior knowledge for distinguishing fine-grained features. Within ZSFDet, Knowledge-Enhanced Feature Synthesizer (KEFS) learns knowledge representation from multiple sources (e.g., ingredients correlation from knowledge graph) via the multi-source graph fusion. Conditioned on the fusion of semantic knowledge representation, the region feature diffusion model in KEFS can generate fine-grained features for training the effective zero-shot detector. Extensive evaluations demonstrate the superior performance of our method ZSFDet on FOWA and the widely-used food dataset UECFOOD-256, with significant improvements by 1.8% and 3.7% ZSD mAP compared with the strong baseline RRFS. Further experiments on PASCAL VOC and MS COCO prove that enhancement of the semantic knowledge can also improve the performance on general ZSD. Code and dataset are available at https://github.com/LanceZPF/KEFS.

link: http://dx.doi.org/10.1109/TIP.2024.3360899

Overview of the L3DAS23 Challenge on Audio-Visual Extended Reality

Christian Marinoni, Riccardo Fosco Gramaccioni, Changan Chen, Aurelio Uncini, Danilo Comminiello

The primary goal of the L3DAS23 Signal Processing Grand Challenge at ICASSP 2023 is to promote and support collaborative research on machine learning for 3D audio signal processing, with a specific emphasis on 3D speech enhancement and 3D Sound Event Localization and Detection in Extended Reality applications. As part of our latest competition, we provide a brand-new dataset, which maintains the same general characteristics of the L3DAS21 and L3DAS22 datasets, but with first-order Ambisonics recordings from multiple reverberant simulated environments. Moreover, we start exploring an audio-visual scenario by providing images of these environments, as perceived by the different microphone positions and orientations. We also propose updated baseline models for both tasks that can now support audio-image couples as input and a supporting API to replicate our results. Finally, we present the results of the participants. Further details about the challenge are available at https://www.l3das.com/icassp2023.

link: http://arxiv.org/abs/2402.09245v1

Who Plays First? Optimizing the Order of Play in Stackelberg Games with Many Robots

Haimin Hu, Gabriele Dragotto, Zixu Zhang, Kaiqu Liang, Bartolomeo Stellato, Jaime F. Fisac

We consider the multi-agent spatial navigation problem of computing the socially optimal order of play, i.e., the sequence in which the agents commit to their decisions, and its associated equilibrium in an N-player Stackelberg trajectory game. We model this problem as a mixed-integer optimization problem over the space of all possible Stackelberg games associated with the order of play's permutations. To solve the problem, we introduce Branch and Play (B&P;), an efficient and exact algorithm that provably converges to a socially optimal order of play and its Stackelberg equilibrium. As a subroutine for B&P;, we employ and extend sequential trajectory planning, i.e., a popular multi-agent control approach, to scalably compute valid local Stackelberg equilibria for any given order of play. We demonstrate the practical utility of B&P; to coordinate air traffic control, swarm formation, and delivery vehicle fleets. We find that B&P; consistently outperforms various baselines, and computes the socially optimal equilibrium.

link: http://arxiv.org/abs/2402.09246v1