Fri 2024.03.29

Secure Aggregation is Not Private Against Membership Inference Attacks

Khac-Hoang Ngo, Johan Östman, Giuseppe Durisi, Alexandre Graell i Amat

Secure aggregation (SecAgg) is a commonly-used privacy-enhancing mechanism in federated learning, affording the server access only to the aggregate of model updates while safeguarding the confidentiality of individual updates. Despite widespread claims regarding SecAgg's privacy-preserving capabilities, a formal analysis of its privacy is lacking, making such presumptions unjustified. In this paper, we delve into the privacy implications of SecAgg by treating it as a local differential privacy (LDP) mechanism for each local update. We design a simple attack wherein an adversarial server seeks to discern which update vector a client submitted, out of two possible ones, in a single training round of federated learning under SecAgg. By conducting privacy auditing, we assess the success probability of this attack and quantify the LDP guarantees provided by SecAgg. Our numerical results unveil that, contrary to prevailing claims, SecAgg offers weak privacy against membership inference attacks even in a single training round. Indeed, it is difficult to hide a local update by adding other independent local updates when the updates are of high dimension. Our findings underscore the imperative for additional privacy-enhancing mechanisms, such as noise injection, in federated learning.

link: http://arxiv.org/abs/2403.17775v1

Exploring the Boundaries of Ambient Awareness in Twitter

Pablo Sanchez-Martin, Sonja Utz, Isabel Valera

Ambient awareness refers to the ability of social media users to obtain knowledge about who knows what (i.e., users' expertise) in their network, by simply being exposed to other users' content (e.g, tweets on Twitter). Previous work, based on user surveys, reveals that individuals self-report ambient awareness only for parts of their networks. However, it is unclear whether it is their limited cognitive capacity or the limited exposure to diagnostic tweets (i.e., online content) that prevents people from developing ambient awareness for their complete network. In this work, we focus on in-wall ambient awareness (IWAA) in Twitter and conduct a two-step data-driven analysis, that allows us to explore to which extent IWAA is likely, or even possible. First, we rely on reactions (e.g., likes), as strong evidence of users being aware of experts in Twitter. Unfortunately, such strong evidence can be only measured for active users, which represent the minority in the network. Thus to study the boundaries of IWAA to a larger extent, in the second part of our analysis, we instead focus on the passive exposure to content generated by other users -- which we refer to as in-wall visibility. This analysis shows that (in line with \citet{levordashka2016ambient}) only for a subset of users IWAA is plausible, while for the majority it is unlikely, if even possible, to develop IWAA. We hope that our methodology paves the way for the emergence of data-driven approaches for the study of ambient awareness.

link: http://arxiv.org/abs/2403.17776v1

Towards a FAIR Documentation of Workflows and Models in Applied Mathematics

Marco Reidelbach, Björn Schembera, Marcus Weber

Modeling-Simulation-Optimization workflows play a fundamental role in applied mathematics. The Mathematical Research Data Initiative, MaRDI, responded to this by developing a FAIR and machine-interpretable template for a comprehensive documentation of such workflows. MaRDMO, a Plugin for the Research Data Management Organiser, enables scientists from diverse fields to document and publish their workflows on the MaRDI Portal seamlessly using the MaRDI template. Central to these workflows are mathematical models. MaRDI addresses them with the MathModDB ontology, offering a structured formal model description. Here, we showcase the interaction between MaRDMO and the MathModDB Knowledge Graph through an algebraic modeling workflow from the Digital Humanities. This demonstration underscores the versatility of both services beyond their original numerical domain.

link: http://arxiv.org/abs/2403.17778v1

GenesisTex: Adapting Image Denoising Diffusion to Texture Space

Chenjian Gao, Boyan Jiang, Xinghui Li, Yingpeng Zhang, Qian Yu

We present GenesisTex, a novel method for synthesizing textures for 3D geometries from text descriptions. GenesisTex adapts the pretrained image diffusion model to texture space by texture space sampling. Specifically, we maintain a latent texture map for each viewpoint, which is updated with predicted noise on the rendering of the corresponding viewpoint. The sampled latent texture maps are then decoded into a final texture map. During the sampling process, we focus on both global and local consistency across multiple viewpoints: global consistency is achieved through the integration of style consistency mechanisms within the noise prediction network, and low-level consistency is achieved by dynamically aligning latent textures. Finally, we apply reference-based inpainting and img2img on denser views for texture refinement. Our approach overcomes the limitations of slow optimization in distillation-based methods and instability in inpainting-based methods. Experiments on meshes from various sources demonstrate that our method surpasses the baseline methods quantitatively and qualitatively.

link: http://arxiv.org/abs/2403.17782v1

SciCapenter: Supporting Caption Composition for Scientific Figures with Machine-Generated Captions and Ratings

Ting-Yao Hsu, Chieh-Yang Huang, Shih-Hong Huang, Ryan Rossi, Sungchul Kim, Tong Yu, C. Lee Giles, Ting-Hao K. Huang

Crafting effective captions for figures is important. Readers heavily depend on these captions to grasp the figure's message. However, despite a well-developed set of AI technologies for figures and captions, these have rarely been tested for usefulness in aiding caption writing. This paper introduces SciCapenter, an interactive system that puts together cutting-edge AI technologies for scientific figure captions to aid caption composition. SciCapenter generates a variety of captions for each figure in a scholarly article, providing scores and a comprehensive checklist to assess caption quality across multiple critical aspects, such as helpfulness, OCR mention, key takeaways, and visual properties reference. Users can directly edit captions in SciCapenter, resubmit for revised evaluations, and iteratively refine them. A user study with Ph.D. students indicates that SciCapenter significantly lowers the cognitive load of caption writing. Participants' feedback further offers valuable design insights for future systems aiming to enhance caption writing.

link: http://dx.doi.org/10.1145/3613905.3650738

Query Refinement for Diverse Top-k Selection

Felix S. Campbell, Alon Silberstein, Julia Stoyanovich, Yuval Moskovitch

Database queries are often used to select and rank items as decision support for many applications. As automated decision-making tools become more prevalent, there is a growing recognition of the need to diversify their outcomes. In this paper, we define and study the problem of modifying the selection conditions of an ORDER BY query so that the result of the modified query closely fits some user-defined notion of diversity while simultaneously maintaining the intent of the original query. We show the hardness of this problem and propose a Mixed Integer Linear Programming (MILP) based solution. We further present optimizations designed to enhance the scalability and applicability of the solution in real-life scenarios. We investigate the performance characteristics of our algorithm and show its efficiency and the usefulness of our optimizations.

link: http://arxiv.org/abs/2403.17786v2

Evaluating the Efficacy of Prompt-Engineered Large Multimodal Models Versus Fine-Tuned Vision Transformers in Image-Based Security Applications

Fouad Trad, Ali Chehab

The success of Large Language Models (LLMs) has led to a parallel rise in the development of Large Multimodal Models (LMMs), such as Gemini-pro, which have begun to transform a variety of applications. These sophisticated multimodal models are designed to interpret and analyze complex data, integrating both textual and visual information on a scale previously unattainable, opening new avenues for a range of applications. This paper investigates the applicability and effectiveness of prompt-engineered Gemini-pro LMMs versus fine-tuned Vision Transformer (ViT) models in addressing critical security challenges. We focus on two distinct tasks: a visually evident task of detecting simple triggers, such as small squares in images, indicative of potential backdoors, and a non-visually evident task of malware classification through visual representations. Our results highlight a significant divergence in performance, with Gemini-pro falling short in accuracy and reliability when compared to fine-tuned ViT models. The ViT models, on the other hand, demonstrate exceptional accuracy, achieving near-perfect performance on both tasks. This study not only showcases the strengths and limitations of prompt-engineered LMMs in cybersecurity applications but also emphasizes the unmatched efficacy of fine-tuned ViT models for precise and dependable tasks.

link: http://arxiv.org/abs/2403.17787v1

Towards 3D Vision with Low-Cost Single-Photon Cameras

Fangzhou Mu, Carter Sifferman, Sacha Jungerman, Yiquan Li, Mark Han, Michael Gleicher, Mohit Gupta, Yin Li

We present a method for reconstructing 3D shape of arbitrary Lambertian objects based on measurements by miniature, energy-efficient, low-cost single-photon cameras. These cameras, operating as time resolved image sensors, illuminate the scene with a very fast pulse of diffuse light and record the shape of that pulse as it returns back from the scene at a high temporal resolution. We propose to model this image formation process, account for its non-idealities, and adapt neural rendering to reconstruct 3D geometry from a set of spatially distributed sensors with known poses. We show that our approach can successfully recover complex 3D shapes from simulated data. We further demonstrate 3D object reconstruction from real-world captures, utilizing measurements from a commodity proximity sensor. Our work draws a connection between image-based modeling and active range scanning and is a step towards 3D vision with single-photon cameras.

link: http://arxiv.org/abs/2403.17801v1

Improving Text-to-Image Consistency via Automatic Prompt Optimization

Oscar Mañas, Pietro Astolfi, Melissa Hall, Candace Ross, Jack Urbanek, Adina Williams, Aishwarya Agrawal, Adriana Romero-Soriano, Michal Drozdzal

Impressive advances in text-to-image (T2I) generative models have yielded a plethora of high performing models which are able to generate aesthetically appealing, photorealistic images. Despite the progress, these models still struggle to produce images that are consistent with the input prompt, oftentimes failing to capture object quantities, relations and attributes properly. Existing solutions to improve prompt-image consistency suffer from the following challenges: (1) they oftentimes require model fine-tuning, (2) they only focus on nearby prompt samples, and (3) they are affected by unfavorable trade-offs among image quality, representation diversity, and prompt-image consistency. In this paper, we address these challenges and introduce a T2I optimization-by-prompting framework, OPT2I, which leverages a large language model (LLM) to improve prompt-image consistency in T2I models. Our framework starts from a user prompt and iteratively generates revised prompts with the goal of maximizing a consistency score. Our extensive validation on two datasets, MSCOCO and PartiPrompts, shows that OPT2I can boost the initial consistency score by up to 24.9% in terms of DSG score while preserving the FID and increasing the recall between generated and real data. Our work paves the way toward building more reliable and robust T2I systems by harnessing the power of LLMs.

link: http://arxiv.org/abs/2403.17804v1

Scenario-Based Curriculum Generation for Multi-Agent Autonomous Driving

Axel Brunnbauer, Luigi Berducci, Peter Priller, Dejan Nickovic, Radu Grosu

The automated generation of diverse and complex training scenarios has been an important ingredient in many complex learning tasks. Especially in real-world application domains, such as autonomous driving, auto-curriculum generation is considered vital for obtaining robust and general policies. However, crafting traffic scenarios with multiple, heterogeneous agents is typically considered as a tedious and time-consuming task, especially in more complex simulation environments. In our work, we introduce MATS-Gym, a Multi-Agent Traffic Scenario framework to train agents in CARLA, a high-fidelity driving simulator. MATS-Gym is a multi-agent training framework for autonomous driving that uses partial scenario specifications to generate traffic scenarios with variable numbers of agents. This paper unifies various existing approaches to traffic scenario description into a single training framework and demonstrates how it can be integrated with techniques from unsupervised environment design to automate the generation of adaptive auto-curricula. The code is available at https://github.com/AutonomousDrivingExaminer/mats-gym.

link: http://arxiv.org/abs/2403.17805v1

Have Faith in Faithfulness: Going Beyond Circuit Overlap When Finding Model Mechanisms

Michael Hanna, Sandro Pezzelle, Yonatan Belinkov

Many recent language model (LM) interpretability studies have adopted the circuits framework, which aims to find the minimal computational subgraph, or circuit, that explains LM behavior on a given task. Most studies determine which edges belong in a LM's circuit by performing causal interventions on each edge independently, but this scales poorly with model size. Edge attribution patching (EAP), gradient-based approximation to interventions, has emerged as a scalable but imperfect solution to this problem. In this paper, we introduce a new method - EAP with integrated gradients (EAP-IG) - that aims to better maintain a core property of circuits: faithfulness. A circuit is faithful if all model edges outside the circuit can be ablated without changing the model's performance on the task; faithfulness is what justifies studying circuits, rather than the full model. Our experiments demonstrate that circuits found using EAP are less faithful than those found using EAP-IG, even though both have high node overlap with circuits found previously using causal interventions. We conclude more generally that when using circuits to compare the mechanisms models use to solve tasks, faithfulness, not overlap, is what should be measured.

link: http://arxiv.org/abs/2403.17806v1

Annotated Biomedical Video Generation using Denoising Diffusion Probabilistic Models and Flow Fields

Rüveyda Yilmaz, Dennis Eschweiler, Johannes Stegmaier

The segmentation and tracking of living cells play a vital role within the biomedical domain, particularly in cancer research, drug development, and developmental biology. These are usually tedious and time-consuming tasks that are traditionally done by biomedical experts. Recently, to automatize these processes, deep learning based segmentation and tracking methods have been proposed. These methods require large-scale datasets and their full potential is constrained by the scarcity of annotated data in the biomedical imaging domain. To address this limitation, we propose Biomedical Video Diffusion Model (BVDM), capable of generating realistic-looking synthetic microscopy videos. Trained only on a single real video, BVDM can generate videos of arbitrary length with pixel-level annotations that can be used for training data-hungry models. It is composed of a denoising diffusion probabilistic model (DDPM) generating high-fidelity synthetic cell microscopy images and a flow prediction model (FPM) predicting the non-rigid transformation between consecutive video frames. During inference, initially, the DDPM imposes realistic cell textures on synthetic cell masks which are generated based on real data statistics. The flow prediction model predicts the flow field between consecutive masks and applies that to the DDPM output from the previous time frame to create the next one while keeping temporal consistency. BVDM outperforms state-of-the-art synthetic live cell microscopy video generation models. Furthermore, we demonstrate that a sufficiently large synthetic dataset enhances the performance of cell segmentation and tracking models compared to using a limited amount of available real data.

link: http://arxiv.org/abs/2403.17808v1

Are Compressed Language Models Less Subgroup Robust?

Leonidas Gee, Andrea Zugarini, Novi Quadrianto

To reduce the inference cost of large language models, model compression is increasingly used to create smaller scalable models. However, little is known about their robustness to minority subgroups defined by the labels and attributes of a dataset. In this paper, we investigate the effects of 18 different compression methods and settings on the subgroup robustness of BERT language models. We show that worst-group performance does not depend on model size alone, but also on the compression method used. Additionally, we find that model compression does not always worsen the performance on minority subgroups. Altogether, our analysis serves to further research into the subgroup robustness of model compression.

link: http://dx.doi.org/10.18653/v1/2023.emnlp-main.983

D-PAD: Deep-Shallow Multi-Frequency Patterns Disentangling for Time Series Forecasting

Xiaobing Yuan, Ling Chen

In time series forecasting, effectively disentangling intricate temporal patterns is crucial. While recent works endeavor to combine decomposition techniques with deep learning, multiple frequencies may still be mixed in the decomposed components, e.g., trend and seasonal. Furthermore, frequency domain analysis methods, e.g., Fourier and wavelet transforms, have limitations in resolution in the time domain and adaptability. In this paper, we propose D-PAD, a deep-shallow multi-frequency patterns disentangling neural network for time series forecasting. Specifically, a multi-component decomposing (MCD) block is introduced to decompose the series into components with different frequency ranges, corresponding to the "shallow" aspect. A decomposition-reconstruction-decomposition (D-R-D) module is proposed to progressively extract the information of frequencies mixed in the components, corresponding to the "deep" aspect. After that, an interaction and fusion (IF) module is used to further analyze the components. Extensive experiments on seven real-world datasets demonstrate that D-PAD achieves the state-of-the-art performance, outperforming the best baseline by an average of 9.48% and 7.15% in MSE and MAE, respectively.

link: http://arxiv.org/abs/2403.17814v1

Graph Language Model (GLM): A new graph-based approach to detect social instabilities

Wallyson Lemes de Oliveira, Vahid Shamsaddini, Ali Ghofrani, Rahul Singh Inda, Jithendra Sai Veeramaneni, Étienne Voutaz

This scientific report presents a novel methodology for the early prediction of important political events using News datasets. The methodology leverages natural language processing, graph theory, clique analysis, and semantic relationships to uncover hidden predictive signals within the data. Initially, we designed a preliminary version of the method and tested it on a few events. This analysis revealed limitations in the initial research phase. We then enhanced the model in two key ways: first, we added a filtration step to only consider politically relevant news before further processing; second, we adjusted the input features to make the alert system more sensitive to significant spikes in the data. After finalizing the improved methodology, we tested it on eleven events including US protests, the Ukraine war, and French protests. Results demonstrate the superiority of our approach compared to baseline methods. Through targeted refinements, our model can now provide earlier and more accurate predictions of major political events based on subtle patterns in news data.

link: http://arxiv.org/abs/2403.17816v1