Sun 2024.04.07

AI and the Problem of Knowledge Collapse

Andrew J. Peterson

While artificial intelligence has the potential to process vast amounts of data, generate new insights, and unlock greater productivity, its widespread adoption may entail unforeseen consequences. We identify conditions under which AI, by reducing the cost of access to certain modes of knowledge, can paradoxically harm public understanding. While large language models are trained on vast amounts of diverse data, they naturally generate output towards the 'center' of the distribution. This is generally useful, but widespread reliance on recursive AI systems could lead to a process we define as "knowledge collapse", and argue this could harm innovation and the richness of human understanding and culture. However, unlike AI models that cannot choose what data they are trained on, humans may strategically seek out diverse forms of knowledge if they perceive them to be worthwhile. To investigate this, we provide a simple model in which a community of learners or innovators choose to use traditional methods or to rely on a discounted AI-assisted process and identify conditions under which knowledge collapse occurs. In our default model, a 20% discount on AI-generated content generates public beliefs 2.3 times further from the truth than when there is no discount. Finally, based on the results, we consider further research directions to counteract such outcomes.

link: http://arxiv.org/abs/2404.03502v1

CountARFactuals -- Generating plausible model-agnostic counterfactual explanations with adversarial random forests

Susanne Dandl, Kristin Blesch, Timo Freiesleben, Gunnar König, Jan Kapar, Bernd Bischl, Marvin Wright

Counterfactual explanations elucidate algorithmic decisions by pointing to scenarios that would have led to an alternative, desired outcome. Giving insight into the model's behavior, they hint users towards possible actions and give grounds for contesting decisions. As a crucial factor in achieving these goals, counterfactuals must be plausible, i.e., describing realistic alternative scenarios within the data manifold. This paper leverages a recently developed generative modeling technique -- adversarial random forests (ARFs) -- to efficiently generate plausible counterfactuals in a model-agnostic way. ARFs can serve as a plausibility measure or directly generate counterfactual explanations. Our ARF-based approach surpasses the limitations of existing methods that aim to generate plausible counterfactual explanations: It is easy to train and computationally highly efficient, handles continuous and categorical data naturally, and allows integrating additional desiderata such as sparsity in a straightforward manner.

link: http://arxiv.org/abs/2404.03506v1

DQ-DETR: DETR with Dynamic Query for Tiny Object Detection

Yi-Xin Huang, Hou-I Liu, Hong-Han Shuai, Wen-Huang Cheng

Despite previous DETR-like methods having performed successfully in generic object detection, tiny object detection is still a challenging task for them since the positional information of object queries is not customized for detecting tiny objects, whose scale is extraordinarily smaller than general objects. Also, DETR-like methods using a fixed number of queries make them unsuitable for aerial datasets, which only contain tiny objects, and the numbers of instances are imbalanced between different images. Thus, we present a simple yet effective model, named DQ-DETR, which consists of three different components: categorical counting module, counting-guided feature enhancement, and dynamic query selection to solve the above-mentioned problems. DQ-DETR uses the prediction and density maps from the categorical counting module to dynamically adjust the number of object queries and improve the positional information of queries. Our model DQ-DETR outperforms previous CNN-based and DETR-like methods, achieving state-of-the-art mAP 30.2% on the AI-TOD-V2 dataset, which mostly consists of tiny objects.

link: http://arxiv.org/abs/2404.03507v1

Learn When (not) to Trust Language Models: A Privacy-Centric Adaptive Model-Aware Approach

Chengkai Huang, Rui Wang, Kaige Xie, Tong Yu, Lina Yao

Retrieval-augmented large language models (LLMs) have been remarkably competent in various NLP tasks. Despite their great success, the knowledge provided by the retrieval process is not always useful for improving the model prediction, since in some samples LLMs may already be quite knowledgeable and thus be able to answer the question correctly without retrieval. Aiming to save the cost of retrieval, previous work has proposed to determine when to do/skip the retrieval in a data-aware manner by analyzing the LLMs' pretraining data. However, these data-aware methods pose privacy risks and memory limitations, especially when requiring access to sensitive or extensive pretraining data. Moreover, these methods offer limited adaptability under fine-tuning or continual learning settings. We hypothesize that token embeddings are able to capture the model's intrinsic knowledge, which offers a safer and more straightforward way to judge the need for retrieval without the privacy risks associated with accessing pre-training data. Moreover, it alleviates the need to retain all the data utilized during model pre-training, necessitating only the upkeep of the token embeddings. Extensive experiments and in-depth analyses demonstrate the superiority of our model-aware approach.

link: http://arxiv.org/abs/2404.03514v1

Model Checking Recursive Probabilistic Programs with Conditioning

Francesco Pontiggia, Ezio Bartocci, Michele Chiari

We address the problem of model checking temporal logic specifications for probabilistic programs with recursive procedures, nested queries, and conditioning expressed with observe statements. We introduce probabilistic Operator Precedence Automata (pOPA), a new class of probabilistic pushdown automata suitable to model constructs and behaviors of probabilistic programs. We develop a model checking algorithm that can verify requirements expressed in a fragment of Precedence Oriented Temporal Logic (POTL$^f_\mathcal{X}$) on a pOPA in single EXPTIME. POTL$^f_\mathcal{X}$ is a temporal logic based on Operator Precedence Languages, which features modalities that interact with the context-free structure of program traces, matching procedure calls with returns or observe statements. We provide the first probabilistic model checking implementation of context-free language properties for probabilistic pushdown systems.

link: http://arxiv.org/abs/2404.03515v1

SDPose: Tokenized Pose Estimation via Circulation-Guide Self-Distillation

Sichen Chen, Yingyi Zhang, Siming Huang, Ran Yi, Ke Fan, Ruixin Zhang, Peixian Chen, Jun Wang, Shouhong Ding, Lizhuang Ma

Recently, transformer-based methods have achieved state-of-the-art prediction quality on human pose estimation(HPE). Nonetheless, most of these top-performing transformer-based models are too computation-consuming and storage-demanding to deploy on edge computing platforms. Those transformer-based models that require fewer resources are prone to under-fitting due to their smaller scale and thus perform notably worse than their larger counterparts. Given this conundrum, we introduce SDPose, a new self-distillation method for improving the performance of small transformer-based models. To mitigate the problem of under-fitting, we design a transformer module named Multi-Cycled Transformer(MCT) based on multiple-cycled forwards to more fully exploit the potential of small model parameters. Further, in order to prevent the additional inference compute-consuming brought by MCT, we introduce a self-distillation scheme, extracting the knowledge from the MCT module to a naive forward model. Specifically, on the MSCOCO validation dataset, SDPose-T obtains 69.7% mAP with 4.4M parameters and 1.8 GFLOPs. Furthermore, SDPose-S-V2 obtains 73.5% mAP on the MSCOCO validation dataset with 6.2M parameters and 4.7 GFLOPs, achieving a new state-of-the-art among predominant tiny neural network methods. Our code is available at https://github.com/MartyrPenink/SDPose.

link: http://arxiv.org/abs/2404.03518v1

Approximate Gradient Coding for Privacy-Flexible Federated Learning with Non-IID Data

Okko Makkonen, Sampo Niemelä, Camilla Hollanti, Serge Kas Hanna

This work focuses on the challenges of non-IID data and stragglers/dropouts in federated learning. We introduce and explore a privacy-flexible paradigm that models parts of the clients' local data as non-private, offering a more versatile and business-oriented perspective on privacy. Within this framework, we propose a data-driven strategy for mitigating the effects of label heterogeneity and client straggling on federated learning. Our solution combines both offline data sharing and approximate gradient coding techniques. Through numerical simulations using the MNIST dataset, we demonstrate that our approach enables achieving a deliberate trade-off between privacy and utility, leading to improved model convergence and accuracy while using an adaptable portion of non-private data.

link: http://arxiv.org/abs/2404.03524v1

HAPNet: Toward Superior RGB-Thermal Scene Parsing via Hybrid, Asymmetric, and Progressive Heterogeneous Feature Fusion

Jiahang Li, Peng Yun, Qijun Chen, Rui Fan

Data-fusion networks have shown significant promise for RGB-thermal scene parsing. However, the majority of existing studies have relied on symmetric duplex encoders for heterogeneous feature extraction and fusion, paying inadequate attention to the inherent differences between RGB and thermal modalities. Recent progress in vision foundation models (VFMs) trained through self-supervision on vast amounts of unlabeled data has proven their ability to extract informative, general-purpose features. However, this potential has yet to be fully leveraged in the domain. In this study, we take one step toward this new research area by exploring a feasible strategy to fully exploit VFM features for RGB-thermal scene parsing. Specifically, we delve deeper into the unique characteristics of RGB and thermal modalities, thereby designing a hybrid, asymmetric encoder that incorporates both a VFM and a convolutional neural network. This design allows for more effective extraction of complementary heterogeneous features, which are subsequently fused in a dual-path, progressive manner. Moreover, we introduce an auxiliary task to further enrich the local semantics of the fused features, thereby improving the overall performance of RGB-thermal scene parsing. Our proposed HAPNet, equipped with all these components, demonstrates superior performance compared to all other state-of-the-art RGB-thermal scene parsing networks, achieving top ranks across three widely used public RGB-thermal scene parsing datasets. We believe this new paradigm has opened up new opportunities for future developments in data-fusion scene parsing approaches.

link: http://arxiv.org/abs/2404.03527v1

BanglaAutoKG: Automatic Bangla Knowledge Graph Construction with Semantic Neural Graph Filtering

Azmine Toushik Wasi, Taki Hasan Rafi, Raima Islam, Dong-Kyu Chae

Knowledge Graphs (KGs) have proven essential in information processing and reasoning applications because they link related entities and give context-rich information, supporting efficient information retrieval and knowledge discovery; presenting information flow in a very effective manner. Despite being widely used globally, Bangla is relatively underrepresented in KGs due to a lack of comprehensive datasets, encoders, NER (named entity recognition) models, POS (part-of-speech) taggers, and lemmatizers, hindering efficient information processing and reasoning applications in the language. Addressing the KG scarcity in Bengali, we propose BanglaAutoKG, a pioneering framework that is able to automatically construct Bengali KGs from any Bangla text. We utilize multilingual LLMs to understand various languages and correlate entities and relations universally. By employing a translation dictionary to identify English equivalents and extracting word features from pre-trained BERT models, we construct the foundational KG. To reduce noise and align word embeddings with our goal, we employ graph-based polynomial filters. Lastly, we implement a GNN-based semantic filter, which elevates contextual understanding and trims unnecessary edges, culminating in the formation of the definitive KG. Empirical findings and case studies demonstrate the universal effectiveness of our model, capable of autonomously constructing semantically enriched KGs from any text.

link: http://arxiv.org/abs/2404.03528v1

COMO: Compact Mapping and Odometry

Eric Dexheimer, Andrew J. Davison

We present COMO, a real-time monocular mapping and odometry system that encodes dense geometry via a compact set of 3D anchor points. Decoding anchor point projections into dense geometry via per-keyframe depth covariance functions guarantees that depth maps are joined together at visible anchor points. The representation enables joint optimization of camera poses and dense geometry, intrinsic 3D consistency, and efficient second-order inference. To maintain a compact yet expressive map, we introduce a frontend that leverages the covariance function for tracking and initializing potentially visually indistinct 3D points across frames. Altogether, we introduce a real-time system capable of estimating accurate poses and consistent geometry.

link: http://arxiv.org/abs/2404.03531v1

Evaluating Generative Language Models in Information Extraction as Subjective Question Correction

Yuchen Fan, Yantao Liu, Zijun Yao, Jifan Yu, Lei Hou, Juanzi Li

Modern Large Language Models (LLMs) have showcased remarkable prowess in various tasks necessitating sophisticated cognitive behaviors. Nevertheless, a paradoxical performance discrepancy is observed, where these models underperform in seemingly elementary tasks like relation extraction and event extraction due to two issues in conventional evaluation. (1) The imprecision of existing evaluation metrics that struggle to effectively gauge semantic consistency between model outputs and ground truth, and (2) The inherent incompleteness of evaluation benchmarks, primarily due to restrictive human annotation schemas, resulting in underestimated LLM performances. Inspired by the principles in subjective question correction, we propose a new evaluation method, SQC-Score. This method innovatively utilizes LLMs, fine-tuned through subjective question correction data, to refine matching between model outputs and golden labels. Additionally, by incorporating a Natural Language Inference (NLI) model, SQC-Score enriches golden labels, addressing benchmark incompleteness by acknowledging correct yet previously omitted answers. Results on three information extraction tasks show that SQC-Score is more preferred by human annotators than the baseline metrics. Utilizing SQC-Score, we conduct a comprehensive evaluation of the state-of-the-art LLMs and provide insights for future research for information extraction. Dataset and associated codes can be accessed at https://github.com/THU-KEG/SQC-Score.

link: http://arxiv.org/abs/2404.03532v1

If It's Not Enough, Make It So: Reducing Authentic Data Demand in Face Recognition through Synthetic Faces

Andrea Atzori, Fadi Boutros, Naser Damer, Gianni Fenu, Mirko Marras

Recent advances in deep face recognition have spurred a growing demand for large, diverse, and manually annotated face datasets. Acquiring authentic, high-quality data for face recognition has proven to be a challenge, primarily due to privacy concerns. Large face datasets are primarily sourced from web-based images, lacking explicit user consent. In this paper, we examine whether and how synthetic face data can be used to train effective face recognition models with reduced reliance on authentic images, thereby mitigating data collection concerns. First, we explored the performance gap among recent state-of-the-art face recognition models, trained with synthetic data only and authentic (scarce) data only. Then, we deepened our analysis by training a state-of-the-art backbone with various combinations of synthetic and authentic data, gaining insights into optimizing the limited use of the latter for verification accuracy. Finally, we assessed the effectiveness of data augmentation approaches on synthetic and authentic data, with the same goal in mind. Our results highlighted the effectiveness of FR trained on combined datasets, particularly when combined with appropriate augmentation techniques.

link: http://arxiv.org/abs/2404.03537v1

Is CLIP the main roadblock for fine-grained open-world perception?

Lorenzo Bianchi, Fabio Carrara, Nicola Messina, Fabrizio Falchi

Modern applications increasingly demand flexible computer vision models that adapt to novel concepts not encountered during training. This necessity is pivotal in emerging domains like extended reality, robotics, and autonomous driving, which require the ability to respond to open-world stimuli. A key ingredient is the ability to identify objects based on free-form textual queries defined at inference time - a task known as open-vocabulary object detection. Multimodal backbones like CLIP are the main enabling technology for current open-world perception solutions. Despite performing well on generic queries, recent studies highlighted limitations on the fine-grained recognition capabilities in open-vocabulary settings - i.e., for distinguishing subtle object features like color, shape, and material. In this paper, we perform a detailed examination of these open-vocabulary object recognition limitations to find the root cause. We evaluate the performance of CLIP, the most commonly used vision-language backbone, against a fine-grained object-matching benchmark, revealing interesting analogies between the limitations of open-vocabulary object detectors and their backbones. Experiments suggest that the lack of fine-grained understanding is caused by the poor separability of object characteristics in the CLIP latent space. Therefore, we try to understand whether fine-grained knowledge is present in CLIP embeddings but not exploited at inference time due, for example, to the unsuitability of the cosine similarity matching function, which may discard important object characteristics. Our preliminary experiments show that simple CLIP latent-space re-projections help separate fine-grained concepts, paving the way towards the development of backbones inherently able to process fine-grained details. The code for reproducing these experiments is available at https://github.com/lorebianchi98/FG-CLIP.

link: http://arxiv.org/abs/2404.03539v1

Segmentation-Guided Knee Radiograph Generation using Conditional Diffusion Models

Siyuan Mei, Fuxin Fan, Fabian Wagner, Mareike Thies, Mingxuan Gu, Yipeng Sun, Andreas Maier

Deep learning-based medical image processing algorithms require representative data during development. In particular, surgical data might be difficult to obtain, and high-quality public datasets are limited. To overcome this limitation and augment datasets, a widely adopted solution is the generation of synthetic images. In this work, we employ conditional diffusion models to generate knee radiographs from contour and bone segmentations. Remarkably, two distinct strategies are presented by incorporating the segmentation as a condition into the sampling and training process, namely, conditional sampling and conditional training. The results demonstrate that both methods can generate realistic images while adhering to the conditioning segmentation. The conditional training method outperforms the conditional sampling method and the conventional U-Net.

link: http://arxiv.org/abs/2404.03541v1

CodeEditorBench: Evaluating Code Editing Capability of Large Language Models

Jiawei Guo, Ziming Li, Xueling Liu, Kaijing Ma, Tianyu Zheng, Zhouliang Yu, Ding Pan, Yizhi LI, Ruibo Liu, Yue Wang, Shuyue Guo, Xingwei Qu, Xiang Yue, Ge Zhang, Wenhu Chen, Jie Fu

Large Language Models (LLMs) for code are rapidly evolving, with code editing emerging as a critical capability. We introduce CodeEditorBench, an evaluation framework designed to rigorously assess the performance of LLMs in code editing tasks, including debugging, translating, polishing, and requirement switching. Unlike existing benchmarks focusing solely on code generation, CodeEditorBench emphasizes real-world scenarios and practical aspects of software development.

We curate diverse coding challenges and scenarios from five sources, covering various programming languages, complexity levels, and editing tasks. Evaluation of 19 LLMs reveals that closed-source models (particularly Gemini-Ultra and GPT-4), outperform open-source models in CodeEditorBench, highlighting differences in model performance based on problem types and prompt sensitivities. CodeEditorBench aims to catalyze advancements in LLMs by providing a robust platform for assessing code editing capabilities. We will release all prompts and datasets to enable the community to expand the dataset and benchmark emerging LLMs. By introducing CodeEditorBench, we contribute to the advancement of LLMs in code editing and provide a valuable resource for researchers and practitioners.

link: http://arxiv.org/abs/2404.03543v1