Thu 2024.04.18

HyperMono: A Monotonicity-aware Approach to Hyper-Relational Knowledge Representation

Zhiwei Hu, Víctor Gutiérrez-Basulto, Zhiliang Xiang, Ru Li, Jeff Z. Pan

In a hyper-relational knowledge graph (HKG), each fact is composed of a main triple associated with attribute-value qualifiers, which express additional factual knowledge. The hyper-relational knowledge graph completion (HKGC) task aims at inferring plausible missing links in a HKG. Most existing approaches to HKGC focus on enhancing the communication between qualifier pairs and main triples, while overlooking two important properties that emerge from the monotonicity of the hyper-relational graphs representation regime. Stage Reasoning allows for a two-step reasoning process, facilitating the integration of coarse-grained inference results derived solely from main triples and fine-grained inference results obtained from hyper-relational facts with qualifiers. In the initial stage, coarse-grained results provide an upper bound for correct predictions, which are subsequently refined in the fine-grained step. More generally, Qualifier Monotonicity implies that by attaching more qualifier pairs to a main triple, we may only narrow down the answer set, but never enlarge it. This paper proposes the HyperMono model for hyper-relational knowledge graph completion, which realizes stage reasoning and qualifier monotonicity. To implement qualifier monotonicity HyperMono resorts to cone embeddings. Experiments on three real-world datasets with three different scenario conditions demonstrate the strong performance of HyperMono when compared to the SoTA.

link: http://arxiv.org/abs/2404.09848v1

Empowering Embodied Visual Tracking with Visual Foundation Models and Offline RL

Fangwei Zhong, Kui Wu, Hai Ci, Churan Wang, Hao Chen

Embodied visual tracking is to follow a target object in dynamic 3D environments using an agent's egocentric vision. This is a vital and challenging skill for embodied agents. However, existing methods suffer from inefficient training and poor generalization. In this paper, we propose a novel framework that combines visual foundation models (VFM) and offline reinforcement learning (offline RL) to empower embodied visual tracking. We use a pre-trained VFM, such as ``Tracking Anything'', to extract semantic segmentation masks with text prompts. We then train a recurrent policy network with offline RL, e.g., Conservative Q-Learning, to learn from the collected demonstrations without online agent-environment interactions. To further improve the robustness and generalization of the policy network, we also introduce a mask re-targeting mechanism and a multi-level data collection strategy. In this way, we can train a robust tracker within an hour on a consumer-level GPU, e.g., Nvidia RTX 3090. Such efficiency is unprecedented for RL-based visual tracking methods. We evaluate our tracker on several high-fidelity environments with challenging situations, such as distraction and occlusion. The results show that our agent outperforms state-of-the-art methods in terms of sample efficiency, robustness to distractors, and generalization to unseen scenarios and targets. We also demonstrate the transferability of the learned tracker from the virtual world to real-world scenarios.

link: http://arxiv.org/abs/2404.09857v1

Unsupervised Federated Optimization at the Edge: D2D-Enabled Learning without Labels

Satyavrat Wagle, Seyyedali Hosseinalipour, Naji Khosravan, Christopher G. Brinton

Federated learning (FL) is a popular solution for distributed machine learning (ML). While FL has traditionally been studied for supervised ML tasks, in many applications, it is impractical to assume availability of labeled data across devices. To this end, we develop Cooperative Federated unsupervised Contrastive Learning ({\tt CF-CL)} to facilitate FL across edge devices with unlabeled datasets. {\tt CF-CL} employs local device cooperation where either explicit (i.e., raw data) or

implicit (i.e., embeddings) information is exchanged through device-to-device (D2D) communications to improve local diversity. Specifically, we introduce a \textit{smart information push-pull} methodology for data/embedding exchange tailored to FL settings with either soft or strict data privacy restrictions. Information sharing is conducted through a probabilistic importance sampling technique at receivers leveraging a carefully crafted reserve dataset provided by transmitters. In the implicit case, embedding exchange is further integrated into the local ML training at the devices via a regularization term incorporated into the contrastive loss, augmented with a dynamic contrastive margin to adjust the volume of latent space explored. Numerical evaluations demonstrate that {\tt CF-CL} leads to alignment of latent spaces learned across devices, results in faster and more efficient global model training, and is effective in extreme non-i.i.d. data distribution settings across devices.

link: http://arxiv.org/abs/2404.09861v1

AI-Driven Statutory Reasoning via Software Engineering Methods

Rohan Padhye

The recent proliferation of generative artificial intelligence (GenAI) technologies such as pre-trained large language models (LLMs) has opened up new frontiers in computational law. An exciting area of development is the use of AI to automate the rule-based reasoning inherent in statutory and contract law. While this form of reasoning has long been studied using classical techniques of natural language processing (NLP) and formal logic, recent solutions increasingly make use of LLMs; though they are far from perfect. The advent of GenAI has made it possible to treat many of these natural language documents essentially as programs that compute a result given some set of facts. As such, it should be possible to understand, debug, maintain, evolve, and fix these documents using well-studied techniques from the field of software engineering. This article introduces several concepts of automated software testing and program analysis that could potentially be useful in computational law when applied to AI-driven analysis of statutes and contracts.

link: http://arxiv.org/abs/2404.09868v1

Table tennis ball spin estimation with an event camera

Thomas Gossard, Julian Krismer, Andreas Ziegler, Jonas Tebbe, Andreas Zell

Spin plays a pivotal role in ball-based sports. Estimating spin becomes a key skill due to its impact on the ball's trajectory and bouncing behavior. Spin cannot be observed directly, making it inherently challenging to estimate. In table tennis, the combination of high velocity and spin renders traditional low frame rate cameras inadequate for quickly and accurately observing the ball's logo to estimate the spin due to the motion blur. Event cameras do not suffer as much from motion blur, thanks to their high temporal resolution. Moreover, the sparse nature of the event stream solves communication bandwidth limitations many frame cameras face. To the best of our knowledge, we present the first method for table tennis spin estimation using an event camera. We use ordinal time surfaces to track the ball and then isolate the events generated by the logo on the ball. Optical flow is then estimated from the extracted events to infer the ball's spin. We achieved a spin magnitude mean error of 10.7 ± 17.3 rps and a spin axis mean error of $32.9 \pm 38.2\deg$ in real time for a flying ball.

link: http://arxiv.org/abs/2404.09870v1

Explainable Online Unsupervised Anomaly Detection for Cyber-Physical Systems via Causal Discovery from Time Series

Daniele Meli

Online unsupervised detection of anomalies is crucial to guarantee the correct operation of cyber-physical systems and the safety of humans interacting with them. State-of-the-art approaches based on deep learning via neural networks achieve outstanding performance at anomaly recognition, evaluating the discrepancy between a normal model of the system (with no anomalies) and the real-time stream of sensor time series. However, large training data and time are typically

required, and explainability is still a challenge to identify the root of the anomaly and implement predictive maintainance. In this paper, we use causal discovery to learn a normal causal graph of the system, and we evaluate the persistency of causal links during real-time acquisition of sensor data to promptly detect anomalies. On two benchmark anomaly detection datasets, we show that our method has higher training efficiency, outperforms the accuracy of state-of-the-art neural architectures and correctly identifies the sources of >10 different anomalies. The code for experimental replication is at http://tinyurl.com/case24causal.

link: http://arxiv.org/abs/2404.09871v1

Conditional Prototype Rectification Prompt Learning

Haoxing Chen, Yaohui Li, Zizheng Huang, Yan Hong, Zhuoer Xu, Zhangxuan Gu, Jun Lan, Huijia Zhu, Weiqiang Wang

Pre-trained large-scale vision-language models (VLMs) have acquired profound understanding of general visual concepts. Recent advancements in efficient transfer learning (ETL) have shown remarkable success in fine-tuning VLMs within the scenario of limited data, introducing only a few parameters to harness task-specific insights from VLMs. Despite significant progress, current leading ETL methods tend to overfit the narrow distributions of base classes seen during training and encounter two primary challenges: (i) only utilizing uni-modal information to modeling task-specific knowledge; and (ii) using costly and time-consuming methods to supplement knowledge. To address these issues, we propose a Conditional Prototype Rectification Prompt Learning (CPR) method to correct the bias of base examples and augment limited data in an effective way. Specifically, we alleviate overfitting on base classes from two aspects. First, each input image acquires knowledge from both textual and visual prototypes, and then generates sample-conditional text tokens. Second, we extract utilizable knowledge from unlabeled data to further refine the prototypes. These two strategies mitigate biases stemming from base classes, yielding a more effective classifier. Extensive experiments on 11 benchmark datasets show that our CPR achieves state-of-the-art performance on both few-shot classification and base-to-new generalization tasks. Our code is avaliable at \url{https://github.com/chenhaoxing/CPR}.

link: http://arxiv.org/abs/2404.09872v1

Synergising Human-like Responses and Machine Intelligence for Planning in Disaster Response

Savvas Papaioannou, Panayiotis Kolios, Christos G. Panayiotou, Marios M. Polycarpou

In the rapidly changing environments of disaster response, planning and decision-making for autonomous agents involve complex and interdependent choices. Although recent advancements have improved traditional artificial intelligence (AI) approaches, they often struggle in such settings, particularly when applied to agents operating outside their well-defined training parameters. To address these challenges, we propose an attention-based cognitive architecture inspired by Dual Process Theory (DPT). This framework integrates, in an online fashion, rapid yet heuristic (human-like) responses (System 1) with the slow but optimized planning capabilities of machine intelligence (System 2). We illustrate how a supervisory controller can dynamically determine in real-time the engagement of either system to optimize mission objectives by assessing their performance across a number of distinct attributes. Evaluated for trajectory planning in dynamic environments, our framework demonstrates that this synergistic integration effectively manages complex tasks by optimizing multiple mission objectives.

link: http://arxiv.org/abs/2404.09877v1

Map-Relative Pose Regression for Visual Re-Localization

Shuai Chen, Tommaso Cavallari, Victor Adrian Prisacariu, Eric Brachmann

Pose regression networks predict the camera pose of a query image relative to a known environment. Within this family of methods, absolute pose regression (APR) has recently shown promising accuracy in the range of a few centimeters in position error. APR networks encode the scene geometry implicitly in their weights. To achieve high accuracy, they require vast amounts of

training data that, realistically, can only be created using novel view synthesis in a days-long process. This process has to be repeated for each new scene again and again. We present a new approach to pose regression, map-relative pose regression (marepo), that satisfies the data hunger of the pose regression network in a scene-agnostic fashion. We condition the pose regressor on a scene-specific map representation such that its pose predictions are relative to the scene map. This allows us to train the pose regressor across hundreds of scenes to learn the generic relation between a scene-specific map representation and the camera pose. Our map-relative pose regressor can be applied to new map representations immediately or after mere minutes of fine-tuning for the highest accuracy. Our approach outperforms previous pose regression methods by far on two public datasets, indoor and outdoor. Code is available: https://nianticlabs.github.io/marepo

link: http://arxiv.org/abs/2404.09884v1

ReffAKD: Resource-efficient Autoencoder-based Knowledge Distillation

Divyang Doshi, Jung-Eun Kim

In this research, we propose an innovative method to boost Knowledge Distillation efficiency without the need for resource-heavy teacher models. Knowledge Distillation trains a smaller ``student'' model with guidance from a larger ``teacher'' model, which is computationally costly. However, the main benefit comes from the soft labels provided by the teacher, helping the student grasp nuanced class similarities. In our work, we propose an efficient method for generating these soft labels, thereby eliminating the need for a large teacher model. We employ a compact autoencoder to extract essential features and calculate similarity scores between different classes. Afterward, we apply the softmax function to these similarity scores to obtain a soft probability vector. This vector serves as valuable guidance during the training of the student model. Our extensive experiments on various datasets, including CIFAR-100, Tiny Imagenet, and Fashion MNIST, demonstrate the superior resource efficiency of our approach compared to traditional knowledge distillation methods that rely on large teacher models. Importantly, our approach consistently achieves similar or even superior performance in terms of model accuracy. We also perform a comparative study with various techniques recently developed for knowledge distillation showing our approach achieves competitive performance with using significantly less resources. We also show that our approach can be easily added to any logit based knowledge distillation method. This research contributes to making knowledge distillation more accessible and cost-effective for practical applications, making it a promising avenue for improving the efficiency of model training. The code for this work is available at, https://github.com/JEKimLab/ReffAKD.

link: http://arxiv.org/abs/2404.09886v1

Is Table Retrieval a Solved Problem? Join-Aware Multi-Table Retrieval

Peter Baile Chen, Yi Zhang, Dan Roth

Retrieving relevant tables containing the necessary information to accurately answer a given question over tables is critical to open-domain question-answering (QA) systems. Previous methods assume the answer to such a question can be found either in a single table or multiple tables identified through question decomposition or rewriting. However, neither of these approaches is sufficient, as many questions require retrieving multiple tables and joining them through a join plan that cannot be discerned from the user query itself. If the join plan is not considered in the retrieval stage, the subsequent steps of reasoning and answering based on those retrieved tables are likely to be incorrect. To address this problem, we introduce a method that uncovers useful join relations for any query and database during table retrieval. We use a novel re-ranking method formulated as a mixed-integer program that considers not only table-query relevance but also table-table relevance that requires inferring join relationships. Our method outperforms the state-of-the-art approaches for table retrieval by up to 9.3% in F1 score and for end-to-end QA by up to 5.4% in accuracy.

link: http://arxiv.org/abs/2404.09889v1

Glitch Tokens in Large Language Models: Categorization Taxonomy and Effective Detection

Yuxi Li, Yi Liu, Gelei Deng, Ying Zhang, Wenjia Song, Ling Shi, Kailong Wang, Yuekang Li, Yang Liu, Haoyu Wang

With the expanding application of Large Language Models (LLMs) in various domains, it becomes imperative to comprehensively investigate their unforeseen behaviors and consequent outcomes. In this study, we introduce and systematically explore the phenomenon of "glitch tokens", which are anomalous tokens produced by established tokenizers and could potentially compromise the models' quality of response. Specifically, we experiment on seven top popular LLMs utilizing three distinct tokenizers and involving a totally of 182,517 tokens. We present categorizations of the identified glitch tokens and symptoms exhibited by LLMs when interacting with glitch tokens. Based on our observation that glitch tokens tend to cluster in the embedding space, we propose GlitchHunter, a novel iterative clustering-based technique, for efficient glitch token detection. The evaluation shows that our approach notably outperforms three baseline methods on eight open-source LLMs. To the best of our knowledge, we present the first comprehensive study on glitch tokens. Our new detection further provides valuable insights into mitigating tokenization-related errors in LLMs.

link: http://arxiv.org/abs/2404.09894v2

Progressive Knowledge Graph Completion

Jiayi Li, Ruilin Luo, Jiaqi Sun, Jing Xiao, Yujiu Yang

Knowledge Graph Completion (KGC) has emerged as a promising solution to address the issue of incompleteness within Knowledge Graphs (KGs). Traditional KGC research primarily centers on triple classification and link prediction. Nevertheless, we contend that these tasks do not align well with real-world scenarios and merely serve as surrogate benchmarks. In this paper, we investigate three crucial processes relevant to real-world construction scenarios: (a) the verification process, which arises from the necessity and limitations of human verifiers; (b) the mining process, which identifies the most promising candidates for verification; and (c) the training process, which harnesses verified data for subsequent utilization; in order to achieve a transition toward more realistic challenges. By integrating these three processes, we introduce the Progressive Knowledge Graph Completion (PKGC) task, which simulates the gradual completion of KGs in real-world scenarios. Furthermore, to expedite PKGC processing, we propose two acceleration modules: Optimized Top-k algorithm and Semantic Validity Filter. These modules significantly enhance the efficiency of the mining procedure. Our experiments demonstrate that performance in link prediction does not accurately reflect performance in PKGC. A more in-depth analysis reveals the key factors influencing the results and provides potential directions for future research.

link: http://arxiv.org/abs/2404.09897v1

Quality of Experience Oriented Cross-layer Optimization for Real-time XR Video Transmission

Guangjin Pan, Shugong Xu, Shunqing Zhang, Xiaojing Chen, Yanzan Sun

Extended reality (XR) is one of the most important applications of beyond 5G and 6G networks. Real-time XR video transmission presents challenges in terms of data rate and delay. In particular, the frame-by-frame transmission mode of XR video makes real-time XR video very sensitive to dynamic network environments. To improve the users' quality of experience (QoE), we design a cross-layer transmission framework for real-time XR video. The proposed framework allows the simple information exchange between the base station (BS) and the XR server, which assists in adaptive bitrate and wireless resource scheduling. We utilize the cross-layer information to formulate the problem of maximizing user QoE by finding the optimal scheduling and bitrate adjustment strategies. To address the issue of mismatched time scales between two strategies, we decouple the original problem and solve them individually using a multi-agent-based approach. Specifically, we propose the multi-step Deep Q-network (MS-DQN) algorithm to obtain a frame-priority-based wireless resource scheduling strategy and then propose the Transformer-based Proximal Policy Optimization (TPPO) algorithm for video bitrate adaptation. The

experimental results show that the TPPO+MS-DQN algorithm proposed in this study can improve the QoE by 3.6% to 37.8%. More specifically, the proposed MS-DQN algorithm enhances the transmission quality by 49.9%-80.2%.

link: http://arxiv.org/abs/2404.09905v1

ChatShop: Interactive Information Seeking with Language Agents

Sanxing Chen, Sam Wiseman, Bhuwan Dhingra

The desire and ability to seek new information strategically are fundamental to human learning but often overlooked in current language agent development. Using a web shopping task as an example, we show that it can be reformulated and solved as a retrieval task without a requirement of interactive information seeking. We then redesign the task to introduce a new role of shopper, serving as a realistically constrained communication channel. The agents in our proposed ChatShop task explore user preferences in open-ended conversation to make informed decisions. Our experiments demonstrate that the proposed task can effectively evaluate the agent's ability to explore and gradually accumulate information through multi-turn interaction. We also show that LLM-simulated shoppers serve as a good proxy to real human shoppers and discover similar error patterns of agents.

link: http://arxiv.org/abs/2404.09911v1