Thu 2024.03.07

Word Importance Explains How Prompts Affect Language Model Outputs

Stefan Hackmann, Haniyeh Mahmoudian, Mark Steadman, Michael Schmidt

The emergence of large language models (LLMs) has revolutionized numerous applications across industries. However, their "black box" nature often hinders the understanding of how they make specific decisions, raising concerns about their transparency, reliability, and ethical use. This study presents a method to improve the explainability of LLMs by varying individual words in prompts to uncover their statistical impact on the model outputs. This approach, inspired by permutation importance for tabular data, masks each word in the system prompt and evaluates its effect on the outputs based on the available text scores aggregated over multiple user inputs. Unlike classical attention, word importance measures the impact of prompt words on arbitrarily-defined text scores, which enables decomposing the importance of words into the specific measures of interest--including bias, reading level, verbosity, etc. This procedure also enables measuring impact when attention weights are not available. To test the fidelity of this approach, we explore the effect of adding different suffixes to multiple different system prompts and comparing subsequent generations with different large language models. Results show that word importance scores are closely related to the expected suffix importances for multiple scoring functions.

link: http://arxiv.org/abs/2403.03028v1

Socratic Reasoning Improves Positive Text Rewriting

Anmol Goel, Nico Daheim, Iryna Gurevych

Reframing a negative into a positive thought is at the crux of several cognitive approaches to mental health and psychotherapy that could be made more accessible by large language model-based solutions. Such reframing is typically non-trivial and requires multiple rationalization steps to uncover the underlying issue of a negative thought and transform it to be more positive. However, this rationalization process is currently neglected by both datasets and models which reframe thoughts in one step. In this work, we address this gap by augmenting open-source datasets for positive text rewriting with synthetically-generated Socratic rationales using a novel framework called \textsc{SocraticReframe}. \textsc{SocraticReframe} uses a sequence of question-answer pairs to rationalize the thought rewriting process. We show that such Socratic rationales significantly improve positive text rewriting for different open-source LLMs according to both automatic and human evaluations guided by criteria from psychotherapy research.

link: http://arxiv.org/abs/2403.03029v1

Unifying Controller Design for Stabilizing Nonlinear Systems with Norm-Bounded Control Inputs

Ming Li, Zhiyong Sun, Siep Weiland

This paper revisits a classical challenge in the design of stabilizing controllers for nonlinear systems with a norm-bounded input constraint. By extending Lin-Sontag's universal formula and introducing a generic (state-dependent) scaling term, a unifying controller design method is proposed. The incorporation of this generic scaling term gives a unified controller and enables the derivation of alternative universal formulas with various favorable properties, which makes it suitable for tailored control designs to meet specific requirements and provides versatility across different control scenarios. Additionally, we present a constructive approach to determine the optimal scaling term, leading to an explicit solution to an optimization problem, named optimization-based universal formula. The resulting controller ensures asymptotic stability, satisfies a norm-bounded input constraint, and optimizes a predefined cost function. Finally, the essential properties of the unified controllers are analyzed, including smoothness, continuity at the origin, stability margin, and inverse optimality. Simulations validate the approach, showcasing its effectiveness in addressing a challenging stabilizing control problem of a nonlinear system.

link: http://arxiv.org/abs/2403.03030v1

Learning to Use Tools via Cooperative and Interactive Agents

Zhengliang Shi, Shen Gao, Xiuyi Chen, Lingyong Yan, Haibo Shi, Dawei Yin, Zhumin Chen, Pengjie Ren, Suzan Verberne, Zhaochun Ren

Tool learning empowers large language models (LLMs) as agents to use external tools to extend their capability. Existing methods employ one single LLM-based agent to iteratively select and execute tools, thereafter incorporating the result into the next action prediction. However, they still suffer from potential performance degradation when addressing complex tasks due to: (1) the limitation of the inherent capability of a single LLM to perform diverse actions, and (2) the struggle to adaptively correct mistakes when the task fails. To mitigate these problems, we propose the ConAgents, a Cooperative and interactive Agents framework, which modularizes the workflow of tool learning into Grounding, Execution, and Observing agents. We also introduce an iterative calibration (IterCali) method, enabling the agents to adapt themselves based on the feedback from the tool environment. Experiments conducted on three datasets demonstrate the superiority of our ConAgents (e.g., 6 point improvement over the SOTA baseline). We further provide fine-granularity analysis for the efficiency and consistency of our framework.

link: http://arxiv.org/abs/2403.03031v1

Mars 2.0: A Toolchain for Modeling, Analysis, Verification and Code Generation of Cyber-Physical Systems

Bohua Zhan, Xiong Xu, Qiang Gao, Zekun Ji, Xiangyu Jin, Shuling Wang, Naijun Zhan

We introduce Mars 2.0 for modeling, analysis, verification and code generation of Cyber-Physical Systems. Mars 2.0 integrates Mars 1.0 with several important extensions and improvements, allowing the design of cyber-physical systems using the combination of AADL and Simulink/Stateflow, which provide a unified graphical framework for modeling the functionality, physicality and architecture of the system to be developed. For a safety-critical system, formal analysis and verification of its combined AADL and Simulink/Stateflow model can be conducted via the following steps. First, the toolchain automatically translates AADL and Simulink/Stateflow models into Hybrid CSP (HCSP), an extension of CSP for formally modeling hybrid systems. Second, the HCSP processes can be simulated using the HCSP simulator, and to complement incomplete simulation, they can be verified using the Hybrid Hoare Logic prover in Isabelle/HOL, as well as the more automated HHLPy prover. Finally, implementations in SystemC or C can be automatically generated from the verified HCSP processes. The transformation from AADL and Simulink/Stateflow to HCSP, and the one from HCSP to SystemC or C, are both guaranteed to be correct with formal proofs. This approach allows model-driven design of safety-critical cyber-physical systems based on graphical and formal models and proven-correct translation procedures. We demonstrate the use of the toolchain on several benchmarks of varying complexity, including several industrial-sized examples.

link: http://arxiv.org/abs/2403.03035v1

A Backpack Full of Skills: Egocentric Video Understanding with Diverse Task Perspectives

Simone Alberto Peirone, Francesca Pistilli, Antonio Alliegro, Giuseppe Averta

Human comprehension of a video stream is naturally broad: in a few instants, we are able to understand what is happening, the relevance and relationship of objects, and forecast what will follow in the near future, everything all at once. We believe that - to effectively transfer such an holistic perception to intelligent machines - an important role is played by learning to correlate concepts and to abstract knowledge coming from different tasks, to synergistically exploit them when learning novel skills. To accomplish this, we seek for a unified approach to video understanding which combines shared temporal modelling of human actions with minimal overhead, to support multiple downstream tasks and enable cooperation when learning novel skills. We then propose EgoPack, a solution that creates a collection of task perspectives that can be

carried across downstream tasks and used as a potential source of additional insights, as a backpack of skills that a robot can carry around and use when needed. We demonstrate the effectiveness and efficiency of our approach on four Ego4D benchmarks, outperforming current state-of-the-art methods.

link: http://arxiv.org/abs/2403.03037v1

Adding Multimodal Capabilities to a Text-only Translation Model

Vipin Vijayan, Braeden Bowen, Scott Grigsby, Timothy Anderson, Jeremy Gwinnup

While most current work in multimodal machine translation (MMT) uses the Multi30k dataset for training and evaluation, we find that the resulting models overfit to the Multi30k dataset to an extreme degree. Consequently, these models perform very badly when evaluated against typical text-only testing sets such as the WMT newstest datasets. In order to perform well on both Multi30k and typical text-only datasets, we use a performant text-only machine translation (MT) model as the starting point of our MMT model. We add vision-text adapter layers connected via gating mechanisms to the MT model, and incrementally transform the MT model into an MMT model by 1) pre-training using vision-based masking of the source text and 2) fine-tuning on Multi30k.

link: http://arxiv.org/abs/2403.03045v1

Neural Codebook Design for Network Beam Management

Ryan M. Dreifuerst, Robert W. Heath Jr

Obtaining accurate and timely channel state information (CSI) is a fundamental challenge for large antenna systems. Mobile systems like 5G use a beam management framework that joins the initial access, beamforming, CSI acquisition, and data transmission. The design of codebooks for these stages, however, is challenging due to their interrelationships, varying array sizes, and site-specific channel and user distributions. Furthermore, beam management is often focused on single-sector operations while ignoring the overarching network- and system-level optimization. In this paper, we proposed an end-to-end learned codebook design algorithm, network beamspace learning (NBL), that captures and optimizes codebooks to mitigate interference while maximizing the achievable performance with extremely large hybrid arrays. The proposed algorithm requires limited shared information yet designs codebooks that outperform traditional codebooks by over 10dB in beam alignment and achieve more than 25% improvements in network spectral efficiency.

link: http://arxiv.org/abs/2403.03053v1

Distributed Policy Gradient for Linear Quadratic Networked Control with Limited Communication Range

Yuzi Yan, Yuan Shen

This paper proposes a scalable distributed policy gradient method and proves its convergence to near-optimal solution in multi-agent linear quadratic networked systems. The agents engage within a specified network under local communication constraints, implying that each agent can only exchange information with a limited number of neighboring agents. On the underlying graph of the network, each agent implements its control input depending on its nearby neighbors' states in the linear quadratic control setting. We show that it is possible to approximate the exact gradient only using local information. Compared with the centralized optimal controller, the performance gap decreases to zero exponentially as the communication and control ranges increase. We also demonstrate how increasing the communication range enhances system stability in the gradient descent process, thereby elucidating a critical trade-off. The simulation results verify our theoretical findings.

link: http://arxiv.org/abs/2403.03055v1

CrackNex: a Few-shot Low-light Crack Segmentation Model Based on Retinex Theory for UAV Inspections

Zhen Yao, Jiawei Xu, Shuhang Hou, Mooi Choo Chuah

Routine visual inspections of concrete structures are imperative for upholding the safety and integrity of critical infrastructure. Such visual inspections sometimes happen under low-light conditions, e.g., checking for bridge health. Crack segmentation under such conditions is challenging due to the poor contrast between cracks and their surroundings. However, most deep learning methods are designed for well-illuminated crack images and hence their performance drops dramatically in low-light scenes. In addition, conventional approaches require many annotated low-light crack images which is time-consuming. In this paper, we address these challenges by proposing CrackNex, a framework that utilizes reflectance information based on Retinex Theory to help the model learn a unified illumination-invariant representation. Furthermore, we utilize few-shot segmentation to solve the inefficient training data problem. In CrackNex, both a support prototype and a reflectance prototype are extracted from the support set. Then, a prototype fusion module is designed to integrate the features from both prototypes. CrackNex outperforms the SOTA methods on multiple datasets. Additionally, we present the first benchmark dataset, LCSD, for low-light crack segmentation. LCSD consists of 102 well-illuminated crack images and 41 low-light crack images. The dataset and code are available at https://github.com/zy1296/CrackNex.

link: http://arxiv.org/abs/2403.03063v1

Enumeration for MSO-Queries on Compressed Trees

Markus Lohrey, Markus L. Schmid

We present a linear preprocessing and output-linear delay enumeration algorithm for MSO-queries over trees that are compressed in the well-established grammar-based framework. Time bounds are measured with respect to the size of the compressed representation of the tree. Our result extends previous work on the enumeration of MSO-queries over uncompressed trees and on the enumeration of document spanners over compressed text documents.

link: http://arxiv.org/abs/2403.03067v1

Improving Variational Autoencoder Estimation from Incomplete Data with Mixture Variational Families

Vaidotas Simkus, Michael U. Gutmann

We consider the task of estimating variational autoencoders (VAEs) when the training data is incomplete. We show that missing data increases the complexity of the model's posterior distribution over the latent variables compared to the fully-observed case. The increased complexity may adversely affect the fit of the model due to a mismatch between the variational and model posterior distributions. We introduce two strategies based on (i) finite variational-mixture and (ii) imputation-based variational-mixture distributions to address the increased posterior complexity. Through a comprehensive evaluation of the proposed approaches, we show that variational mixtures are effective at improving the accuracy of VAE estimation from incomplete data.

link: http://arxiv.org/abs/2403.03069v1

On a Neural Implementation of Brenier's Polar Factorization

Nina Vesseron, Marco Cuturi

In 1991, Brenier proved a theorem that generalizes the QR decomposition for square matrices -- factored as PSD $\times$ unitary -- to any vector field $F:\mathbb{R}^d\rightarrow \mathbb{R}^d$. The theorem, known as the polar factorization theorem, states that any field F can be recovered as the composition of the gradient of a convex function u with a measure-preserving map M, namely $F=\nabla u \circ M$. We propose a practical implementation of this far-reaching theoretical result, and explore possible uses within machine learning. The theorem is closely related to optimal transport (OT) theory, and we borrow from recent advances in the field of neural optimal transport to parameterize the potential u as an input convex neural network. The map M can be either evaluated pointwise using u^*, the convex conjugate of u, through the identity $M=\nabla u^* \circ F$, or learned as an auxiliary network. Because M is, in general, not injective, we consider the additional task of estimating the ill-posed inverse map that can approximate the pre-image measure M^{-1} using a stochastic generator. We illustrate possible applications of

\citeauthor{Brenier1991PolarFA}'s polar factorization to non-convex optimization problems, as well as sampling of densities that are not log-concave.

link: http://arxiv.org/abs/2403.03071v1

Detecting Concrete Visual Tokens for Multimodal Machine Translation

Braeden Bowen, Vipin Vijayan, Scott Grigsby, Timothy Anderson, Jeremy Gwinnup

The challenge of visual grounding and masking in multimodal machine translation (MMT) systems has encouraged varying approaches to the detection and selection of visually-grounded text tokens for masking. We introduce new methods for detection of visually and contextually relevant (concrete) tokens from source sentences, including detection with natural language processing (NLP), detection with object detection, and a joint detection-verification technique. We also introduce new methods for selection of detected tokens, including shortest n tokens, longest n tokens, and all detected concrete tokens. We utilize the GRAM MMT architecture to train models against synthetically collated multimodal datasets of source images with masked sentences, showing performance improvements and improved usage of visual context during translation tasks over the baseline model.

link: http://arxiv.org/abs/2403.03075v1

MiKASA: Multi-Key-Anchor & Scene-Aware Transformer for 3D Visual Grounding

Chun-Peng Chang, Shaoxiang Wang, Alain Pagani, Didier Stricker

3D visual grounding involves matching natural language descriptions with their corresponding objects in 3D spaces. Existing methods often face challenges with accuracy in object recognition and struggle in interpreting complex linguistic queries, particularly with descriptions that involve multiple anchors or are view-dependent. In response, we present the MiKASA (Multi-Key-Anchor Scene-Aware) Transformer. Our novel end-to-end trained model integrates a self-attention-based scene-aware object encoder and an original multi-key-anchor technique, enhancing object recognition accuracy and the understanding of spatial relationships. Furthermore, MiKASA improves the explainability of decision-making, facilitating error diagnosis. Our model achieves the highest overall accuracy in the Referit3D challenge for both the Sr3D and Nr3D datasets, particularly excelling by a large margin in categories that require viewpoint-dependent descriptions. The source code and additional resources for this project are available on GitHub: https://github.com/birdy666/MiKASA-3DVG

link: http://arxiv.org/abs/2403.03077v1