Tue 2024.03.12

LinearAPT: An Adaptive Algorithm for the Fixed-Budget Thresholding Linear Bandit Problem

Yun-Ang Wu, Yun-Da Tsai, Shou-De Lin

In this study, we delve into the Thresholding Linear Bandit (TLB) problem, a nuanced domain within stochastic Multi-Armed Bandit (MAB) problems, focusing on maximizing decision accuracy against a linearly defined threshold under resource constraints. We present LinearAPT, a novel algorithm designed for the fixed budget setting of TLB, providing an efficient solution to optimize sequential decision-making. This algorithm not only offers a theoretical upper bound for estimated loss but also showcases robust performance on both synthetic and real-world datasets. Our contributions highlight the adaptability, simplicity, and computational efficiency of LinearAPT, making it a valuable addition to the toolkit for addressing complex sequential decision-making challenges.

link: http://arxiv.org/abs/2403.06230v1

Finding Visual Saliency in Continuous Spike Stream

Lin Zhu, Xianzhang Chen, Xiao Wang, Hua Huang

As a bio-inspired vision sensor, the spike camera emulates the operational principles of the fovea, a compact retinal region, by employing spike discharges to encode the accumulation of per-pixel luminance intensity. Leveraging its high temporal resolution and bio-inspired neuromorphic design, the spike camera holds significant promise for advancing computer vision applications. Saliency detection mimics the behavior of human beings and captures the most salient region from the scenes. In this paper, we investigate the visual saliency in the continuous spike stream for the first time. To effectively process the binary spike stream, we propose a Recurrent Spiking Transformer (RST) framework, which is based on a full spiking neural network. Our framework enables the extraction of spatio-temporal features from the continuous spatio-temporal spike stream while maintaining low power consumption. To facilitate the training and validation of our proposed model, we build a comprehensive real-world spike-based visual saliency dataset, enriched with numerous light conditions. Extensive experiments demonstrate the superior performance of our Recurrent Spiking Transformer framework in comparison to other spike neural network-based methods. Our framework exhibits a substantial margin of improvement in capturing and highlighting visual saliency in the spike stream, which not only provides a new perspective for spike-based saliency segmentation but also shows a new paradigm for full SNN-based transformer models. The code and dataset are available at \url{https://github.com/BIT-Vision/SVS}.

link: http://arxiv.org/abs/2403.06233v1

Probabilistic Neural Circuits

Pedro Zuidberg Dos Martires

Probabilistic circuits (PCs) have gained prominence in recent years as a versatile framework for discussing probabilistic models that support tractable queries and are yet expressive enough to model complex probability distributions. Nevertheless, tractability comes at a cost: PCs are less expressive than neural networks. In this paper we introduce probabilistic neural circuits (PNCs), which strike a balance between PCs and neural nets in terms of tractability and expressive power. Theoretically, we show that PNCs can be interpreted as deep mixtures of Bayesian networks. Experimentally, we demonstrate that PNCs constitute powerful function approximators.

link: http://arxiv.org/abs/2403.06235v1

Cooperative Classification and Rationalization for Graph Generalization

Linan Yue, Qi Liu, Ye Liu, Weibo Gao, Fangzhou Yao, Wenfeng Li

Graph Neural Networks (GNNs) have achieved impressive results in graph classification tasks, but they struggle to generalize effectively when faced with out-of-distribution (OOD) data. Several approaches have been proposed to address this problem. Among them, one solution is to diversify training distributions in vanilla classification by modifying the data environment, yet accessing the environment information is complex. Besides, another promising approach involves rationalization, extracting invariant rationales for predictions. However, extracting rationales is difficult due to limited learning signals, resulting in less accurate rationales and diminished predictions. To address these challenges, in this paper, we propose a Cooperative Classification and Rationalization (C2R) method, consisting of the classification and the rationalization module. Specifically, we first assume that multiple environments are available in the classification module. Then, we introduce diverse training distributions using an environment-conditional generative network, enabling robust graph representations. Meanwhile, the rationalization module employs a separator to identify relevant rationale subgraphs while the remaining non-rationale subgraphs are de-correlated with labels. Next, we align graph representations from the classification module with rationale subgraph representations using the knowledge distillation methods, enhancing the learning signal for rationales. Finally, we infer multiple environments by gathering non-rationale representations and incorporate them into the classification module for cooperative learning. Extensive experimental results on both benchmarks and synthetic datasets demonstrate the effectiveness of C2R. Code is available at https://github.com/yuelinan/Codes-of-C2R.

link: http://arxiv.org/abs/2403.06239v1

COVID-19 Computer-aided Diagnosis through AI-assisted CT Imaging Analysis: Deploying a Medical AI System

Demetris Gerogiannis, Anastasios Arsenos, Dimitrios Kollias, Dimitris Nikitopoulos, Stefanos Kollias

Computer-aided diagnosis (CAD) systems stand out as potent aids for physicians in identifying the novel Coronavirus Disease 2019 (COVID-19) through medical imaging modalities. In this paper, we showcase the integration and reliable and fast deployment of a state-of-the-art AI system designed to automatically analyze CT images, offering infection probability for the swift detection of COVID-19. The suggested system, comprising both classification and segmentation components, is anticipated to reduce physicians' detection time and enhance the overall efficiency of COVID-19 detection. We successfully surmounted various challenges, such as data discrepancy and anonymisation, testing the time-effectiveness of the model, and data security, enabling reliable and scalable deployment of the system on both cloud and edge environments. Additionally, our AI system assigns a probability of infection to each 3D CT scan and enhances explainability through anchor set similarity, facilitating timely confirmation and segregation of infected patients by physicians.

link: http://arxiv.org/abs/2403.06242v1

BlazeBVD: Make Scale-Time Equalization Great Again for Blind Video Deflickering

Xinmin Qiu, Congying Han, Zicheng Zhang, Bonan Li, Tiande Guo, Pingyu Wang, Xuecheng Nie

Developing blind video deflickering (BVD) algorithms to enhance video temporal consistency, is gaining importance amid the flourish of image processing and video generation. However, the intricate nature of video data complicates the training of deep learning methods, leading to high resource consumption and instability, notably under severe lighting flicker. This underscores the critical need for a compact representation beyond pixel values to advance BVD research and applications. Inspired by the classic scale-time equalization (STE), our work introduces the histogram-assisted solution, called BlazeBVD, for high-fidelity and rapid BVD. Compared with STE, which directly corrects pixel values by temporally smoothing color histograms, BlazeBVD leverages smoothed illumination histograms within STE filtering to ease the challenge of learning temporal data using neural networks. In technique, BlazeBVD begins by condensing pixel values into illumination histograms that precisely capture flickering and local exposure variations. These histograms are then smoothed to produce singular frames set, filtered illumination maps, and exposure maps. Resorting to these deflickering priors, BlazeBVD utilizes a 2D network to restore

faithful and consistent texture impacted by lighting changes or localized exposure issues. BlazeBVD also incorporates a lightweight 3D network to amend slight temporal inconsistencies, avoiding the resource consumption issue. Comprehensive experiments on synthetic, real-world and generated videos, showcase the superior qualitative and quantitative results of BlazeBVD, achieving inference speeds up to 10x faster than state-of-the-arts.

link: http://arxiv.org/abs/2403.06243v1

Text-Guided Variational Image Generation for Industrial Anomaly Detection and Segmentation

Mingyu Lee, Jongwon Choi

We propose a text-guided variational image generation method to address the challenge of getting clean data for anomaly detection in industrial manufacturing. Our method utilizes text information about the target object, learned from extensive text library documents, to generate non-defective data images resembling the input image. The proposed framework ensures that the generated non-defective images align with anticipated distributions derived from textual and image-based knowledge, ensuring stability and generality. Experimental results demonstrate the effectiveness of our approach, surpassing previous methods even with limited non-defective data. Our approach is validated through generalization tests across four baseline models and three distinct datasets. We present an additional analysis to enhance the effectiveness of anomaly detection models by utilizing the generated images.

link: http://arxiv.org/abs/2403.06247v1

No Language is an Island: Unifying Chinese and English in Financial Large Language Models, Instruction Data, and Benchmarks

Gang Hu, Ke Qin, Chenhan Yuan, Min Peng, Alejandro Lopez-Lira, Benyou Wang, Sophia Ananiadou, Wanlong Yu, Jimin Huang, Qianqian Xie

While the progression of Large Language Models (LLMs) has notably propelled financial analysis, their application has largely been confined to singular language realms, leaving untapped the potential of bilingual Chinese-English capacity. To bridge this chasm, we introduce ICE-PIXIU, seamlessly amalgamating the ICE-INTENT model and ICE-FLARE benchmark for bilingual financial analysis. ICE-PIXIU uniquely integrates a spectrum of Chinese tasks, alongside translated and original English datasets, enriching the breadth and depth of bilingual financial modeling. It provides unrestricted access to diverse model variants, a substantial compilation of diverse cross-lingual and multi-modal instruction data, and an evaluation benchmark with expert annotations, comprising 10 NLP tasks, 20 bilingual specific tasks, totaling 1,185k datasets. Our thorough evaluation emphasizes the advantages of incorporating these bilingual datasets, especially in translation tasks and utilizing original English data, enhancing both linguistic flexibility and analytical acuity in financial contexts. Notably, ICE-INTENT distinguishes itself by showcasing significant enhancements over conventional LLMs and existing financial LLMs in bilingual milieus, underscoring the profound impact of robust bilingual data on the accuracy and efficacy of financial NLP.

link: http://arxiv.org/abs/2403.06249v1

Poly Kernel Inception Network for Remote Sensing Detection

Xinhao Cai, Qiuxia Lai, Yuwei Wang, Wenguan Wang, Zeren Sun, Yazhou Yao

Object detection in remote sensing images (RSIs) often suffers from several increasing challenges, including the large variation in object scales and the diverse-ranging context. Prior methods tried to address these challenges by expanding the spatial receptive field of the backbone, either through large-kernel convolution or dilated convolution. However, the former typically introduces considerable background noise, while the latter risks generating overly sparse feature representations. In this paper, we introduce the Poly Kernel Inception Network (PKINet) to handle the above challenges. PKINet employs multi-scale convolution kernels without dilation to extract object features of varying scales and capture local context. In addition, a Context Anchor Attention

(CAA) module is introduced in parallel to capture long-range contextual information. These two components work jointly to advance the performance of PKINet on four challenging remote sensing detection benchmarks, namely DOTA-v1.0, DOTA-v1.5, HRSC2016, and DIOR-R.

link: http://arxiv.org/abs/2403.06258v1

Editing Conceptual Knowledge for Large Language Models

Xiaohan Wang, Shengyu Mao, Ningyu Zhang, Shumin Deng, Yunzhi Yao, Yue Shen, Lei Liang, Jinjie Gu, Huajun Chen

Recently, there has been a growing interest in knowledge editing for Large Language Models (LLMs). Current approaches and evaluations merely explore the instance-level editing, while whether LLMs possess the capability to modify concepts remains unclear. This paper pioneers the investigation of editing conceptual knowledge for LLMs, by constructing a novel benchmark dataset ConceptEdit and establishing a suite of new metrics for evaluation. The experimental results reveal that, although existing editing methods can efficiently modify concept-level definition to some extent, they also have the potential to distort the related instantial knowledge in LLMs, leading to poor performance. We anticipate this can inspire further progress in better understanding LLMs. Our project homepage is available at https://zjunlp.github.io/project/ConceptEdit.

link: http://arxiv.org/abs/2403.06259v1

SCORE: Self-supervised Correspondence Fine-tuning for Improved Content Representations

Amit Meghanani, Thomas Hain

There is a growing interest in cost-effective self-supervised fine-tuning (SSFT) of self-supervised learning (SSL)-based speech models to obtain task-specific representations. These task-specific representations are used for robust performance on various downstream tasks by fine-tuning on the labelled data. This work presents a cost-effective SSFT method named Self-supervised Correspondence (SCORE) fine-tuning to adapt the SSL speech representations for content-related tasks. The proposed method uses a correspondence training strategy, aiming to learn similar representations from perturbed speech and original speech. Commonly used data augmentation techniques for content-related tasks (ASR) are applied to obtain perturbed speech. SCORE fine-tuned HuBERT outperforms the vanilla HuBERT on SUPERB benchmark with only a few hours of fine-tuning (< 5 hrs) on a single GPU for automatic speech recognition, phoneme recognition, and query-by-example tasks, with relative improvements of 1.09%, 3.58%, and 12.65%, respectively. SCORE provides competitive results with the recently proposed SSFT method SPIN, using only 1/3 of the processed speech compared to SPIN.

link: http://arxiv.org/abs/2403.06260v1

Unpacking Tokenization: Evaluating Text Compression and its Correlation with Model Performance

Omer Goldman, Avi Caciularu, Matan Eyal, Kris Cao, Idan Szpektor, Reut Tsarfaty

Despite it being the cornerstone of BPE, the most common tokenization algorithm, the importance of compression in the tokenization process is still unclear. In this paper, we argue for the theoretical importance of compression, that can be viewed as 0-gram language modeling where equal probability is assigned to all tokens. We also demonstrate the empirical importance of compression for downstream success of pre-trained language models. We control the compression ability of several BPE tokenizers by varying the amount of documents available during their training: from 1 million documents to a character-based tokenizer equivalent to no training data at all. We then pre-train English language models based on those tokenizers and fine-tune them over several tasks. We show that there is a correlation between tokenizers' compression and models' downstream performance, suggesting that compression is a reliable intrinsic indicator of tokenization quality. These correlations are more pronounced for generation tasks (over classification) or for smaller models (over large ones). We replicated a representative part of our experiments on Turkish and found similar results, confirming that our results hold for languages

with typological characteristics dissimilar to English. We conclude that building better compressing tokenizers is a fruitful avenue for further research and for improving overall model performance.

link: http://arxiv.org/abs/2403.06265v1

FARPLS: A Feature-Augmented Robot Trajectory Preference Labeling System to Assist Human Labelers' Preference Elicitation

Hanfang Lyu, Yuanchen Bai, Xin Liang, Ujaan Das, Chuhan Shi, Leiliang Gong, Yingchi Li, Mingfei Sun, Ming Ge, Xiaojuan Ma

Preference-based learning aims to align robot task objectives with human values. One of the most common methods to infer human preferences is by pairwise comparisons of robot task trajectories. Traditional comparison-based preference labeling systems seldom support labelers to digest and identify critical differences between complex trajectories recorded in videos. Our formative study (N = 12) suggests that individuals may overlook non-salient task features and establish biased preference criteria during their preference elicitation process because of partial observations. In addition, they may experience mental fatigue when given many pairs to compare, causing their label quality to deteriorate. To mitigate these issues, we propose FARPLS, a Feature-Augmented Robot trajectory Preference Labeling System. FARPLS highlights potential outliers in a wide variety of task features that matter to humans and extracts the corresponding video keyframes for easy review and comparison. It also dynamically adjusts the labeling order according to users' familiarities, difficulties of the trajectory pair, and level of disagreements. At the same time, the system monitors labelers' consistency and provides feedback on labeling progress to keep labelers engaged. A between-subjects study (N = 42, 105 pairs of robot pick-and-place trajectories per person) shows that FARPLS can help users establish preference criteria more easily and notice more relevant details in the presented trajectories than the conventional interface. FARPLS also improves labeling consistency and engagement, mitigating challenges in preference elicitation without raising cognitive loads significantly

link: http://dx.doi.org/10.1145/3640543.3645145

Physics-Guided Abnormal Trajectory Gap Detection

Arun Sharma, Shashi Shekhar

Given trajectories with gaps (i.e., missing data), we investigate algorithms to identify abnormal gaps in trajectories which occur when a given moving object did not report its location, but other moving objects in the same geographic region periodically did. The problem is important due to its societal applications, such as improving maritime safety and regulatory enforcement for global security concerns such as illegal fishing, illegal oil transfers, and trans-shipments. The problem is challenging due to the difficulty of bounding the possible locations of the moving object during a trajectory gap, and the very high computational cost of detecting gaps in such a large volume of location data. The current literature on anomalous trajectory detection assumes linear interpolation within gaps, which may not be able to detect abnormal gaps since objects within a given region may have traveled away from their shortest path. In preliminary work, we introduced an abnormal gap measure that uses a classical space-time prism model to bound an object's possible movement during the trajectory gap and provided a scalable memoized gap detection algorithm (Memo-AGD). In this paper, we propose a Space Time-Aware Gap Detection (STAGD) approach to leverage space-time indexing and merging of trajectory gaps. We also incorporate a Dynamic Region Merge-based (DRM) approach to efficiently compute gap abnormality scores. We provide theoretical proofs that both algorithms are correct and complete and also provide analysis of asymptotic time complexity. Experimental results on synthetic and real-world maritime trajectory data show that the proposed approach substantially improves computation time over the baseline technique.

link: http://arxiv.org/abs/2403.06268v1

FastVideoEdit: Leveraging Consistency Models for Efficient Text-to-Video Editing

Youyuan Zhang, Xuan Ju, James J. Clark

Diffusion models have demonstrated remarkable capabilities in text-to-image and text-to-video generation, opening up possibilities for video editing based on textual input. However, the computational cost associated with sequential sampling in diffusion models poses challenges for efficient video editing. Existing approaches relying on image generation models for video editing suffer from time-consuming one-shot fine-tuning, additional condition extraction, or DDIM inversion, making real-time applications impractical. In this work, we propose FastVideoEdit, an efficient zero-shot video editing approach inspired by Consistency Models (CMs). By leveraging the self-consistency property of CMs, we eliminate the need for time-consuming inversion or additional condition extraction, reducing editing time. Our method enables direct mapping from source video to target video with strong preservation ability utilizing a special variance schedule. This results in improved speed advantages, as fewer sampling steps can be used while maintaining comparable generation quality. Experimental results validate the state-of-the-art performance and speed advantages of FastVideoEdit across evaluation metrics encompassing editing speed, temporal consistency, and text-video alignment.

link: http://arxiv.org/abs/2403.06269v1