Tue 2024.05.14

Lasso Ridge based XGBoost and Deep_LSTM Help Tennis Players Perform better

Wankang Zhai, Yuhan Wang

Understanding the dynamics of momentum and game fluctuation in tennis matches is cru-cial for predicting match outcomes and enhancing player performance. In this study, we present a comprehensive analysis of these factors using a dataset from the 2023 Wimbledon final. Ini-tially, we develop a sliding-window-based scoring model to assess player performance, ac-counting for the influence of serving dominance through a serve decay factor. Additionally, we introduce a novel approach, Lasso-Ridge-based XGBoost, to quantify momentum effects, lev-eraging the predictive power of XGBoost while mitigating overfitting through regularization. Through experimentation, we achieve an accuracy of 94% in predicting match outcomes, iden-tifying key factors influencing winning rates. Subsequently, we propose a Derivative of the winning rate algorithm to quantify game fluctuation, employing an LSTM_Deep model to pre-dict fluctuation scores. Our model effectively captures temporal correlations in momentum fea-tures, yielding mean squared errors ranging from 0.036 to 0.064. Furthermore, we explore me-ta-learning using MAML to transfer our model to predict outcomes in ping-pong matches, though results indicate a comparative performance decline. Our findings provide valuable in-sights into momentum dynamics and game fluctuation, offering implications for sports analytics and player training strategies.

link: http://arxiv.org/abs/2405.07030v1

Global Motion Understanding in Large-Scale Video Object Segmentation

Volodymyr Fedynyak, Yaroslav Romanus, Oles Dobosevych, Igor Babin, Roman Riazantsev

In this paper, we show that transferring knowledge from other domains of video understanding combined with large-scale learning can improve robustness of Video Object Segmentation (VOS) under complex circumstances. Namely, we focus on integrating scene global motion knowledge to improve large-scale semi-supervised Video Object Segmentation. Prior works on VOS mostly rely on direct comparison of semantic and contextual features to perform dense matching between current and past frames, passing over actual motion structure. On the other hand, Optical Flow Estimation task aims to approximate the scene motion field, exposing global motion patterns which are typically undiscoverable during all pairs similarity search. We present WarpFormer, an architecture for semi-supervised Video Object Segmentation that exploits existing knowledge in motion understanding to conduct smoother propagation and more accurate matching. Our framework employs a generic pretrained Optical Flow Estimation network whose prediction is used to warp both past frames and instance segmentation masks to the current frame domain. Consequently, warped segmentation masks are refined and fused together aiming to inpaint occluded regions and eliminate artifacts caused by flow field imperfects. Additionally, we employ novel large-scale MOSE 2023 dataset to train model on various complex scenarios. Our method demonstrates strong performance on DAVIS 2016/2017 validation (93.0% and 85.9%), DAVIS 2017 test-dev (80.6%) and YouTube-VOS 2019 validation (83.8%) that is competitive with alternative state-of-the-art methods while using much simpler memory mechanism and instance understanding logic.

link: http://arxiv.org/abs/2405.07031v1

A Performance Analysis Modeling Framework for Extended Reality Applications in Edge-Assisted Wireless Networks

Anik Mallik, Jiang Xie, Zhu Han

Extended reality (XR) is at the center of attraction in the research community due to the emergence of augmented, mixed, and virtual reality applications. The performance of such applications needs to be uptight to maintain the requirements of latency, energy consumption, and freshness of data. Therefore, a comprehensive performance analysis model is required to assess the effectiveness of an XR application but is challenging to design due to the dependence of the performance metrics on several difficult-to-model parameters, such as computing resources and hardware utilization of XR and edge devices, which are controlled by both their operating systems and the application itself. Moreover, the heterogeneity in devices and wireless access networks brings additional challenges in modeling. In this paper, we propose a novel modeling framework for performance analysis of XR applications considering edge-assisted wireless networks and validate the model with experimental data collected from testbeds designed specifically for XR applications. In addition, we present the challenges associated with performance analysis modeling and present methods to overcome them in detail. Finally, the performance evaluation shows that the proposed analytical model can analyze XR applications' performance with high accuracy compared to the state-of-the-art analytical models.

link: http://arxiv.org/abs/2405.07033v1

A Turkish Educational Crossword Puzzle

Kamyar Zeinalipour, Yusuf Gökberk Kepti■, Marco Maggini, Leonardo Rigutini, Marco Gori

This paper introduces the first Turkish crossword puzzle generator designed to leverage the capabilities of large language models (LLMs) for educational purposes. In this work, we introduced two specially created datasets: one with over 180,000 unique answer-clue pairs for generating relevant clues from the given answer, and another with over 35,000 samples containing text, answer, category, and clue data, aimed at producing clues for specific texts and keywords within certain categories. Beyond entertainment, this generator emerges as an interactive educational tool that enhances memory, vocabulary, and problem-solving skills. It's a notable step in AI-enhanced education, merging game-like engagement with learning for Turkish and setting new standards for interactive, intelligent learning tools in Turkish.

link: http://arxiv.org/abs/2405.07035v1

Conformal Online Auction Design

Jiale Han, Xiaowu Dai

This paper proposes the conformal online auction design (COAD), a novel mechanism for maximizing revenue in online auctions by quantifying the uncertainty in bidders' values without relying on assumptions about value distributions. COAD incorporates both the bidder and item features and leverages historical data to provide an incentive-compatible mechanism for online auctions. Unlike traditional methods for online auctions, COAD employs a distribution-free, prediction interval-based approach using conformal prediction techniques. This novel approach ensures that the expected revenue from our mechanism can achieve at least a constant fraction of the revenue generated by the optimal mechanism. Additionally, COAD admits the use of a broad array of modern machine-learning methods, including random forests, kernel methods, and deep neural nets, for predicting bidders' values. It ensures revenue performance under any finite sample of historical data. Moreover, COAD introduces bidder-specific reserve prices based on the lower confidence bounds of bidders' valuations, which is different from the uniform reserve prices commonly used in the literature. We validate our theoretical predictions through extensive simulations and a real-data application. All code for using COAD and reproducing results is made available on GitHub.

link: http://arxiv.org/abs/2405.07038v1

Multi-agent Traffic Prediction via Denoised Endpoint Distribution

Yao Liu, Ruoyu Wang, Yuanjiang Cao, Quan Z. Sheng, Lina Yao

The exploration of high-speed movement by robots or road traffic agents is crucial for autonomous driving and navigation. Trajectory prediction at high speeds requires considering historical features and interactions with surrounding entities, a complexity not as pronounced in lower-speed environments. Prior methods have assessed the spatio-temporal dynamics of agents but often neglected intrinsic intent and uncertainty, thereby limiting their effectiveness. We present the Denoised Endpoint Distribution model for trajectory prediction, which distinctively models agents' spatio-temporal features alongside their intrinsic intentions and uncertainties. By employing

Diffusion and Transformer models to focus on agent endpoints rather than entire trajectories, our approach significantly reduces model complexity and enhances performance through endpoint information. Our experiments on open datasets, coupled with comparison and ablation studies, demonstrate our model's efficacy and the importance of its components. This approach advances trajectory prediction in high-speed scenarios and lays groundwork for future developments.

link: http://arxiv.org/abs/2405.07041v1

Semantic Guided Large Scale Factor Remote Sensing Image Super-resolution with Generative Diffusion Prior

Ce Wang, Wanjie Sun

Remote sensing images captured by different platforms exhibit significant disparities in spatial resolution. Large scale factor super-resolution (SR) algorithms are vital for maximizing the utilization of low-resolution (LR) satellite data captured from orbit. However, existing methods confront challenges in recovering SR images with clear textures and correct ground objects. We introduce a novel framework, the Semantic Guided Diffusion Model (SGDM), designed for large scale factor remote sensing image super-resolution. The framework exploits a pre-trained generative model as a prior to generate perceptually plausible SR images. We further enhance the reconstruction by incorporating vector maps, which carry structural and semantic cues. Moreover, pixel-level inconsistencies in paired remote sensing images, stemming from sensor-specific imaging characteristics, may hinder the convergence of the model and diversity in generated results. To address this problem, we propose to extract the sensor-specific imaging characteristics and model the distribution of them, allowing diverse SR images generation based on imaging characteristics provided by reference images or sampled from the imaging characteristic probability distributions. To validate and evaluate our approach, we create the Cross-Modal Super-Resolution Dataset (CMSRD). Qualitative and quantitative experiments on CMSRD showcase the superiority and broad applicability of our method. Experimental results on downstream vision tasks also demonstrate the utilitarian of the generated SR images. The dataset and code will be publicly available at https://github.com/wwangcece/SGDM

link: http://arxiv.org/abs/2405.07044v1

Predictive Modeling in the Reservoir Kernel Motif Space

Peter Tino, Robert Simon Fong, Roberto Fabio Leonarduzzi

This work proposes a time series prediction method based on the kernel view of linear reservoirs. In particular, the time series motifs of the reservoir kernel are used as representational basis on which general readouts are constructed. We provide a geometric interpretation of our approach shedding light on how our approach is related to the core reservoir models and in what way the two approaches differ. Empirical experiments then compare predictive performances of our suggested model with those of recent state-of-art transformer based models, as well as the established recurrent network model - LSTM. The experiments are performed on both univariate and multivariate time series and with a variety of prediction horizons. Rather surprisingly we show that even when linear readout is employed, our method has the capacity to outperform transformer models on univariate time series and attain competitive results on multivariate benchmark datasets. We conclude that simple models with easily controllable capacity but capturing enough memory and subsequence structure can outperform potentially over-complicated deep learning models. This does not mean that reservoir motif based models are preferable to other more complex alternatives - rather, when introducing a new complex time series model one should employ as a sanity check simple, but potentially powerful alternatives/baselines such as reservoir models or the models introduced here.

link: http://arxiv.org/abs/2405.07045v1

Retrieval Enhanced Zero-Shot Video Captioning

Yunchuan Ma, Laiyun Qing, Guorong Li, Yuankai Qi, Quan Z. Sheng, Qingming Huang

Despite the significant progress of fully-supervised video captioning, zero-shot methods remain much less explored. In this paper, we propose to take advantage of existing pre-trained large-scale vision and language models to directly generate captions with test time adaptation. Specifically, we bridge video and text using three key models: a general video understanding model XCLIP, a general image understanding model CLIP, and a text generation model GPT-2, due to their source-code availability. The main challenge is how to enable the text generation model to be sufficiently aware of the content in a given video so as to generate corresponding captions. To address this problem, we propose using learnable tokens as a communication medium between frozen GPT-2 and frozen XCLIP as well as frozen CLIP. Differing from the conventional way to train these tokens with training data, we update these tokens with pseudo-targets of the inference data under several carefully crafted loss functions which enable the tokens to absorb video information catered for GPT-2. This procedure can be done in just a few iterations (we use 16 iterations in the experiments) and does not require ground truth data. Extensive experimental results on three widely used datasets, MSR-VTT, MSVD, and VATEX, show 4% to 20% improvements in terms of the main metric CIDEr compared to the existing state-of-the-art methods.

link: http://arxiv.org/abs/2405.07046v1

Unsupervised Density Neural Representation for CT Metal Artifact Reduction

Qing Wu, Xu Guo, Lixuan Chen, Dongming He, Hongjiang Wei, Xudong Wang, S. Kevin Zhou, Yifeng Zhang, Jingyi Yu, Yuyao Zhang

Emerging unsupervised reconstruction techniques based on implicit neural representation (INR), such as NeRP, CoIL, and SCOPE, have shown unique capabilities in CT linear inverse imaging. In this work, we propose a novel unsupervised density neural representation (Diner) to tackle the challenging problem of CT metal artifacts when scanned objects contain metals. The drastic variation of linear attenuation coefficients (LACs) of metals over X-ray spectra leads to a nonlinear beam hardening effect (BHE) in CT measurements. Recovering CT images from metal-affected measurements therefore poses a complicated nonlinear inverse problem. Existing metal artifact reduction (MAR) techniques mostly formulate the MAR as an image inpainting task, which ignores the energy-induced BHE and produces suboptimal performance. Instead, our Diner introduces an energy-dependent polychromatic CT forward model to the INR framework, addressing the nonlinear nature of the MAR problem. Specifically, we decompose the energy-dependent LACs into energy-independent densities and energy-dependent mass attenuation coefficients (MACs) by fully considering the physical model of X-ray absorption. Using the densities as pivot variables and the MACs as known prior knowledge, the LACs can be accurately reconstructed from the raw measurements. Technically, we represent the unknown density map as an implicit function of coordinates. Combined with a novel differentiable forward model simulating the physical acquisition from the densities to the measurements, our Diner optimizes a multi-layer perception network to approximate the implicit function by minimizing predicted errors between the estimated and real measurements. Experimental results on simulated and real datasets confirm the superiority of our unsupervised Diner against popular supervised techniques in MAR performance and robustness.

link: http://arxiv.org/abs/2405.07047v1

Length-Aware Multi-Kernel Transformer for Long Document Classification

Guangzeng Han, Jack Tsao, Xiaolei Huang

Lengthy documents pose a unique challenge to neural language models due to substantial memory consumption. While existing state-of-the-art (SOTA) models segment long texts into equal-length snippets (e.g., 128 tokens per snippet) or deploy sparse attention networks, these methods have new challenges of context fragmentation and generalizability due to sentence boundaries and varying text lengths. For example, our empirical analysis has shown that SOTA models consistently overfit one set of lengthy documents (e.g., 2000 tokens) while performing worse on texts with other lengths (e.g., 1000 or 4000). In this study, we propose a Length-Aware Multi-Kernel Transformer (LAMKIT) to address the new challenges for the long document classification. LAMKIT encodes lengthy documents by diverse transformer-based kernels for bridging context boundaries and vectorizes text length by the kernels to promote model robustness over varying document lengths.

Experiments on five standard benchmarks from health and law domains show LAMKIT outperforms SOTA models up to an absolute 10.9% improvement. We conduct extensive ablation analyses to examine model robustness and effectiveness over varying document lengths.

link: http://arxiv.org/abs/2405.07052v1

LLMs and the Future of Chip Design: Unveiling Security Risks and Building Trust

Zeng Wang, Lilas Alrahis, Likhitha Mankali, Johann Knechtel, Ozgur Sinanoglu

Chip design is about to be revolutionized by the integration of large language, multimodal, and circuit models (collectively LxMs). While exploring this exciting frontier with tremendous potential, the community must also carefully consider the related security risks and the need for building trust into using LxMs for chip design. First, we review the recent surge of using LxMs for chip design in general. We cover state-of-the-art works for the automation of hardware description language code generation and for scripting and guidance of essential but cumbersome tasks for electronic design automation tools, e.g., design-space exploration, tuning, or designer training. Second, we raise and provide initial answers to novel research questions on critical issues for security and trustworthiness of LxM-powered chip design from both the attack and defense perspectives.

link: http://arxiv.org/abs/2405.07061v1

Learning Flame Evolution Operator under Hybrid Darrieus Landau and Diffusive Thermal Instability

Rixin Yu, Erdzan Hodzic, Karl-Johan Nogenmyr

Recent advancements in the integration of artificial intelligence (AI) and machine learning (ML) with physical sciences have led to significant progress in addressing complex phenomena governed by nonlinear partial differential equations (PDE). This paper explores the application of novel operator learning methodologies to unravel the intricate dynamics of flame instability, particularly focusing on hybrid instabilities arising from the coexistence of Darrieus-Landau (DL) and Diffusive-Thermal (DT) mechanisms. Training datasets encompass a wide range of parameter configurations, enabling the learning of parametric solution advancement operators using techniques such as parametric Fourier Neural Operator (pFNO), and parametric convolutional neural networks (pCNN). Results demonstrate the efficacy of these methods in accurately predicting short-term and long-term flame evolution across diverse parameter regimes, capturing the characteristic behaviors of pure and blended instabilities. Comparative analyses reveal pFNO as the most accurate model for learning short-term solutions, while all models exhibit robust performance in capturing the nuanced dynamics of flame evolution. This research contributes to the development of robust modeling frameworks for understanding and controlling complex physical processes governed by nonlinear PDE.

link: http://arxiv.org/abs/2405.07067v1

Catastrophe Insurance: An Adaptive Robust Optimization Approach

Dimitris Bertsimas, Cynthia Zeng

The escalating frequency and severity of natural disasters, exacerbated by climate change, underscore the critical role of insurance in facilitating recovery and promoting investments in risk reduction. This work introduces a novel Adaptive Robust Optimization (ARO) framework tailored for the calculation of catastrophe insurance premiums, with a case study applied to the United States National Flood Insurance Program (NFIP). To the best of our knowledge, it is the first time an ARO approach has been applied to for disaster insurance pricing. Our methodology is designed to protect against both historical and emerging risks, the latter predicted by machine learning models, thus directly incorporating amplified risks induced by climate change. Using the US flood insurance data as a case study, optimization models demonstrate effectiveness in covering losses and produce surpluses, with a smooth balance transition through parameter fine-tuning. Among tested optimization models, results show ARO models with conservative parameter values achieving low number of insolvent states with the least insurance premium charged. Overall, optimization frameworks offer versatility and generalizability, making it adaptable to a variety of natural disaster scenarios, such as wildfires, droughts, etc. This work not only advances the field of insurance premium modeling but also serves as a vital tool for policymakers and stakeholders in building resilience to the growing risks of natural catastrophes.

link: http://arxiv.org/abs/2405.07068v1

Decoding Cognitive Health Using Machine Learning: A Comprehensive Evaluation for Diagnosis of Significant Memory Concern

M. Sajid, Rahul Sharma, Iman Beheshti, M. Tanveer

The timely identification of significant memory concern (SMC) is crucial for proactive cognitive health management, especially in an aging population. Detecting SMC early enables timely intervention and personalized care, potentially slowing cognitive disorder progression. This study presents a state-of-the-art review followed by a comprehensive evaluation of machine learning models within the randomized neural networks (RNNs) and hyperplane-based classifiers (HbCs) family to investigate SMC diagnosis thoroughly. Utilizing the Alzheimer's Disease Neuroimaging Initiative 2 (ADNI2) dataset, 111 individuals with SMC and 111 healthy older adults are analyzed based on T1W magnetic resonance imaging (MRI) scans, extracting rich features. This analysis is based on baseline structural MRI (sMRI) scans, extracting rich features from gray matter (GM), white matter (WM), Jacobian determinant (JD), and cortical thickness (CT) measurements. In RNNs, deep random vector functional link (dRVFL) and ensemble dRVFL (edRVFL) emerge as the best classifiers in terms of performance metrics in the identification of SMC. In HbCs, Kernelized pinball general twin support vector machine (Pin-GTSVM-K) excels in CT and WM features, whereas Linear Pin-GTSVM (Pin-GTSVM-L) and Linear intuitionistic fuzzy TSVM (IFTSVM-L) performs well in the JD and GM features sets, respectively. This comprehensive evaluation emphasizes the critical role of feature selection and model choice in attaining an effective classifier for SMC diagnosis. The inclusion of statistical analyses further reinforces the credibility of the results, affirming the rigor of this analysis. The performance measures exhibit the suitability of this framework in aiding researchers with the automated and accurate assessment of SMC. The source codes of the algorithms and datasets used in this study are available at https://github.com/mtanveer1/SMC.

link: http://arxiv.org/abs/2405.07070v1