Thu 2024.02.29

An Empirical Study of the Generalization Ability of Lidar 3D Object Detectors to Unseen Domains

George Eskandar, Chongzhe Zhang, Abhishek Kaushik, Karim Guirguis, Mohamed Sayed, Bin Yang

3D Object Detectors (3D-OD) are crucial for understanding the environment in many robotic tasks, especially autonomous driving. Including 3D information via Lidar sensors improves accuracy greatly. However, such detectors perform poorly on domains they were not trained on, i.e. different locations, sensors, weather, etc., limiting their reliability in safety-critical applications. There exist methods to adapt 3D-ODs to these domains; however, these methods treat 3D-ODs as a black box, neglecting underlying architectural decisions and source-domain training strategies. Instead, we dive deep into the details of 3D-ODs, focusing our efforts on fundamental factors that influence robustness prior to domain adaptation. We systematically investigate four design choices (and the interplay between them) often overlooked in 3D-OD robustness and domain adaptation: architecture, voxel encoding, data augmentations, and anchor strategies. We assess their impact on the robustness of nine state-of-the-art 3D-ODs across six benchmarks encompassing three types of domain gaps - sensor type, weather, and location. Our main findings are: (1) transformer backbones with local point features are more robust than 3D CNNs, (2) test-time anchor size adjustment is crucial for adaptation across geographical locations, significantly boosting scores without retraining, (3) source-domain augmentations allow the model to generalize to low-resolution sensors, and (4) surprisingly, robustness to bad weather is improved when training directly on more clean weather data than on training with bad weather data. We outline our main conclusions and findings to provide practical guidance on developing more robust 3D-ODs.

link: http://arxiv.org/abs/2402.17562v1

Structure-Guided Adversarial Training of Diffusion Models

Ling Yang, Haotian Qian, Zhilong Zhang, Jingwei Liu, Bin Cui

Diffusion models have demonstrated exceptional efficacy in various generative applications. While existing models focus on minimizing a weighted sum of denoising score matching losses for data distribution modeling, their training primarily emphasizes instance-level optimization, overlooking valuable structural information within each mini-batch, indicative of pair-wise relationships among samples. To address this limitation, we introduce Structure-guided Adversarial training of Diffusion Models (SADM). In this pioneering approach, we compel the model to learn manifold structures between samples in each training batch. To ensure the model captures authentic manifold structures in the data distribution, we advocate adversarial training of the diffusion generator against a novel structure discriminator in a minimax game, distinguishing real manifold structures from the generated ones. SADM substantially improves existing diffusion transformers (DiT) and outperforms existing methods in image generation and cross-domain fine-tuning tasks across 12 datasets, establishing a new state-of-the-art FID of 1.58 and 2.11 on ImageNet for class-conditional image generation at resolutions of 256x256 and 512x512, respectively.

link: http://arxiv.org/abs/2402.17563v1

Unleashing the Potential of Large Language Models as Prompt Optimizers: An Analogical Analysis with Gradient-based Model Optimizers

Xinyu Tang, Xiaolei Wang, Wayne Xin Zhao, Siyuan Lu, Yaliang Li, Ji-Rong Wen

Automatic prompt optimization is an important approach to improving the performance of large language models (LLMs). Recent research demonstrates the potential of using LLMs as prompt optimizers, which can generate improved task prompts via iterative refinement. In this paper, we propose a novel perspective to investigate the design of LLM-based prompt optimizers, by drawing an analogy with gradient-based model optimizers. To connect these two approaches, we identify two pivotal factors in model parameter learning: update direction and update method. Focused on

the two aspects, we borrow the theoretical framework and learning methods from gradient-based optimization to design improved strategies for LLM-based prompt optimizers. By systematically analyzing a rich set of improvement strategies, we further develop a capable Gradient-inspired LLM-based Prompt Optimizer called GPO. At each step, it first retrieves relevant prompts from the optimization trajectory as the update direction. Then, it utilizes the generation-based refinement strategy to perform the update, while controlling the edit distance through a cosine-based decay strategy. Extensive experiments demonstrate the effectiveness and efficiency of GPO. In particular, GPO brings an additional improvement of up to 56.8% on Big-Bench Hard and 55.3% on MMLU compared to baseline methods.

link: http://arxiv.org/abs/2402.17564v1

Sparse Variational Contaminated Noise Gaussian Process Regression for Forecasting Geomagnetic Perturbations

Daniel Iong, Matthew McAnear, Yuezhou Qu, Shasha Zou, Gabor Toth, Yang Chen

Gaussian Processes (GP) have become popular machine learning methods for kernel based learning on datasets with complicated covariance structures. In this paper, we present a novel extension to the GP framework using a contaminated normal likelihood function to better account for heteroscedastic variance and outlier noise. We propose a scalable inference algorithm based on the Sparse Variational Gaussian Process (SVGP) method for fitting sparse Gaussian process regression models with contaminated normal noise on large datasets. We examine an application to geomagnetic ground perturbations, where the state-of-art prediction model is based on neural networks. We show that our approach yields shorter predictions intervals for similar coverage and accuracy when compared to an artificial dense neural network baseline.

link: http://arxiv.org/abs/2402.17570v2

HBF MU-MIMO with Interference-Aware Beam Pair Link Allocation for Beyond-5G mm-Wave Networks

Aleksandar Ichkov, Alexander Wietfeld, Marina Petrova, Ljiljana Simi■

Hybrid beamforming (HBF) multi-user multiple-input multiple-output (MU-MIMO) is a key technology for unlocking the directional millimeter-wave (mm-wave) nature for spatial multiplexing beyond current codebook-based 5G-NR networks. In order to suppress co-scheduled users' interference, HBF MU-MIMO is predicated on having sufficient radio frequency chains and accurate channel state information (CSI), which can otherwise lead to performance losses due to imperfect interference cancellation. In this work, we propose IABA, a 5G-NR standard-compliant beam pair link (BPL) allocation scheme for mitigating spatial interference in practical HBF MU-MIMO networks. IABA solves the network sum throughput optimization via either a distributed or a centralized BPL allocation using dedicated CSI reference signals for candidate BPL monitoring. We present a comprehensive study of practical multi-cell mm-wave networks and demonstrate that HBF MU-MIMO without interference-aware BPL allocation experiences strong residual interference which limits the achievable network performance. Our results show that IABA offers significant performance gains over the default interference-agnostic 5G-NR BPL allocation, and even allows HBF MU-MIMO to outperform the fully digital MU-MIMO baseline, by facilitating allocation of secondary BPLs other than the strongest BPL found during initial access. We further demonstrate the scalability of IABA with increased gNB antennas and densification for beyond-5G mm-wave networks.

link: http://arxiv.org/abs/2402.17573v1

Agent-Pro: Learning to Evolve via Policy-Level Reflection and Optimization

Wenqi Zhang, Ke Tang, Hai Wu, Mengna Wang, Yongliang Shen, Guiyang Hou, Zeqi Tan, Peng Li, Yueting Zhuang, Weiming Lu

Large Language Models exhibit robust problem-solving capabilities for diverse tasks. However, most LLM-based agents are designed as specific task solvers with sophisticated prompt engineering, rather than agents capable of learning and evolving through interactions. These task

solvers necessitate manually crafted prompts to inform task rules and regulate LLM behaviors, inherently incapacitating to address complex dynamic scenarios e.g., large interactive games. In light of this, we propose Agent-Pro: an LLM-based Agent with Policy-level Reflection and Optimization that can learn a wealth of expertise from interactive experiences and progressively elevate its behavioral policy. Specifically, it involves a dynamic belief generation and reflection process for policy evolution. Rather than action-level reflection, Agent-Pro iteratively reflects on past trajectories and beliefs, fine-tuning its irrational beliefs for a better policy. Moreover, a depth-first search is employed for policy optimization, ensuring continual enhancement in policy payoffs. Agent-Pro is evaluated across two games: Blackjack and Texas Hold'em, outperforming vanilla LLM and specialized models. Our results show Agent-Pro can learn and evolve in complex and dynamic scenes, which also benefits numerous LLM-based applications.

link: http://arxiv.org/abs/2402.17574v1

FaultProfIT: Hierarchical Fault Profiling of Incident Tickets in Large-scale Cloud Systems

Junjie Huang, Jinyang Liu, Zhuangbin Chen, Zhihan Jiang, Yichen LI, Jiazhen Gu, Cong Feng, Zengyin Yang, Yongqiang Yang, Michael R. Lyu

Postmortem analysis is essential in the management of incidents within cloud systems, which provides valuable insights to improve system's reliability and robustness. At CloudA, fault pattern profiling is performed during the postmortem phase, which involves the classification of incidents' faults into unique categories, referred to as fault pattern. By aggregating and analyzing these fault patterns, engineers can discern common faults, vulnerable components and emerging fault trends. However, this process is currently conducted by manual labeling, which has inherent drawbacks. On the one hand, the sheer volume of incidents means only the most severe ones are analyzed, causing a skewed overview of fault patterns. On the other hand, the complexity of the task demands extensive domain knowledge, which leads to errors and inconsistencies. To address these limitations, we propose an automated approach, named FaultProfIT, for Fault pattern Profiling of Incident Tickets. It leverages hierarchy-guided contrastive learning to train a hierarchy-aware incident encoder and predicts fault patterns with enhanced incident representations. We evaluate FaultProfIT using the production incidents from CloudA. The results demonstrate that FaultProfIT outperforms state-of-the-art methods. Our ablation study and analysis also verify the effectiveness of hierarchy-guided contrastive learning. Additionally, we have deployed FaultProfIT at CloudA for six months. To date, FaultProfIT has analyzed 10,000+ incidents from 30+ cloud services, successfully revealing several fault trends that have informed system improvements.

link: http://dx.doi.org/10.1145/3639477.3639754

PLReMix: Combating Noisy Labels with Pseudo-Label Relaxed Contrastive Representation Learning

Xiaoyu Liu, Beitong Zhou, Cheng Cheng

Recently, the application of Contrastive Representation Learning (CRL) in learning with noisy labels (LNL) has shown promising advancements due to its remarkable ability to learn well-distributed representations for better distinguishing noisy labels. However, CRL is mainly used as a pre-training technique, leading to a complicated multi-stage training pipeline. We also observed that trivially combining CRL with supervised LNL methods decreases performance. Using different images from the same class as negative pairs in CRL creates optimization conflicts between CRL and the supervised loss. To address these two issues, we propose an end-to-end PLReMix framework that avoids the complicated pipeline by introducing a Pseudo-Label Relaxed (PLR) contrastive loss to alleviate the conflicts between losses. This PLR loss constructs a reliable negative set of each sample by filtering out its inappropriate negative pairs that overlap at the top k indices of prediction probabilities, leading to more compact semantic clusters than vanilla CRL. Furthermore, a two-dimensional Gaussian Mixture Model (GMM) is adopted to distinguish clean and noisy samples by leveraging semantic information and model outputs simultaneously, which is expanded on the previously widely used one-dimensional form. The PLR loss and a semi-supervised loss are simultaneously applied to train on the GMM divided clean and noisy

samples. Experiments on multiple benchmark datasets demonstrate the effectiveness of the proposed method. Our proposed PLR loss is scalable, which can be easily integrated into other LNL methods and boost their performance. Codes will be available.

link: http://arxiv.org/abs/2402.17589v1

Implicit Regularization via Spectral Neural Networks and Non-linear Matrix Sensing

Hong T. M. Chu, Subhro Ghosh, Chi Thanh Lam, Soumendu Sundar Mukherjee

The phenomenon of implicit regularization has attracted interest in recent years as a fundamental aspect of the remarkable generalizing ability of neural networks. In a nutshell, it entails that gradient descent dynamics in many neural nets, even without any explicit regularizer in the loss function, converges to the solution of a regularized learning problem. However, known results attempting to theoretically explain this phenomenon focus overwhelmingly on the setting of linear neural nets, and the simplicity of the linear structure is particularly crucial to existing arguments. In this paper, we explore this problem in the context of more realistic neural networks with a general class of non-linear activation functions, and rigorously demonstrate the implicit regularization phenomenon for such networks in the setting of matrix sensing problems, together with rigorous rate guarantees that ensure exponentially fast convergence of gradient descent.In this vein, we contribute a network architecture called Spectral Neural Networks (abbrv. SNN) that is particularly suitable for matrix learning problems. Conceptually, this entails coordinatizing the space of matrices by their singular values and singular vectors, as opposed to by their entries, a potentially fruitful perspective for matrix learning. We demonstrate that the SNN architecture is inherently much more amenable to theoretical analysis than vanilla neural nets and confirm its effectiveness in the context of matrix sensing, via both mathematical guarantees and empirical investigations. We believe that the SNN architecture has the potential to be of wide applicability in a broad class of matrix learning scenarios.

link: http://arxiv.org/abs/2402.17595v1

Advancing sleep detection by modelling weak label sets: A novel weakly supervised learning approach

Matthias Boeker, Vajira Thambawita, Michael Riegler, Pål Halvorsen, Hugo L. Hammer

Understanding sleep and activity patterns plays a crucial role in physical and mental health. This study introduces a novel approach for sleep detection using weakly supervised learning for scenarios where reliable ground truth labels are unavailable. The proposed method relies on a set of weak labels, derived from the predictions generated by conventional sleep detection algorithms. Introducing a novel approach, we suggest a novel generalised non-linear statistical model in which the number of weak sleep labels is modelled as outcome of a binomial distribution. The probability of sleep in the binomial distribution is linked to the outcomes of neural networks trained to detect sleep based on actigraphy. We show that maximizing the likelihood function of the model, is equivalent to minimizing the soft cross-entropy loss. Additionally, we explored the use of the Brier score as a loss function for weak labels. The efficacy of the suggested modelling framework was demonstrated using the Multi-Ethnic Study of Atherosclerosis dataset. A \gls{lstm} trained on the soft cross-entropy outperformed conventional sleep detection algorithms, other neural network architectures and loss functions in accuracy and model calibration. This research not only advances sleep detection techniques in scenarios where ground truth data is scarce but also contributes to the broader field of weakly supervised learning by introducing innovative approach in modelling sets of weak labels.

link: http://arxiv.org/abs/2402.17601v1

Learning Topological Representations with Bidirectional Graph Attention Network for Solving Job Shop Scheduling Problem

Cong Zhang, Zhiguang Cao, Yaoxin Wu, Wen Song, Jing Sun

Existing learning-based methods for solving job shop scheduling problem (JSSP) usually use off-the-shelf GNN models tailored to undirected graphs and neglect the rich and meaningful

topological structures of disjunctive graphs (DGs). This paper proposes the topology-aware bidirectional graph attention network (TBGAT), a novel GNN architecture based on the attention mechanism, to embed the DG for solving JSSP in a local search framework. Specifically, TBGAT embeds the DG from a forward and a backward view, respectively, where the messages are propagated by following the different topologies of the views and aggregated via graph attention. Then, we propose a novel operator based on the message-passing mechanism to calculate the forward and backward topological sorts of the DG, which are the features for characterizing the topological structures and exploited by our model. In addition, we theoretically and experimentally show that TBGAT has linear computational complexity to the number of jobs and machines, respectively, which strengthens the practical value of our method. Besides, extensive experiments on five synthetic datasets and seven classic benchmarks show that TBGAT achieves new SOTA results by outperforming a wide range of neural methods by a large margin.

link: http://arxiv.org/abs/2402.17606v1

Linguistic Knowledge Can Enhance Encoder-Decoder Models (If You Let It)

Alessio Miaschi, Felice Dell'Orletta, Giulia Venturi

In this paper, we explore the impact of augmenting pre-trained Encoder-Decoder models, specifically T5, with linguistic knowledge for the prediction of a target task. In particular, we investigate whether fine-tuning a T5 model on an intermediate task that predicts structural linguistic properties of sentences modifies its performance in the target task of predicting sentence-level complexity. Our study encompasses diverse experiments conducted on Italian and English datasets, employing both monolingual and multilingual T5 models at various sizes. Results obtained for both languages and in cross-lingual configurations show that linguistically motivated intermediate fine-tuning has generally a positive impact on target task performance, especially when applied to smaller models and in scenarios with limited data availability.

link: http://arxiv.org/abs/2402.17608v1

A Large-scale Evaluation of Pretraining Paradigms for the Detection of Defects in Electroluminescence Solar Cell Images

David Torpey, Lawrence Pratt, Richard Klein

Pretraining has been shown to improve performance in many domains, including semantic segmentation, especially in domains with limited labelled data. In this work, we perform a large-scale evaluation and benchmarking of various pretraining methods for Solar Cell Defect Detection (SCDD) in electroluminescence images, a field with limited labelled datasets. We cover supervised training with semantic segmentation, semi-supervised learning, and two self-supervised techniques. We also experiment with both in-distribution and out-of-distribution (OOD) pretraining and observe how this affects downstream performance. The results suggest that supervised training on a large OOD dataset (COCO), self-supervised pretraining on a large OOD dataset (ImageNet), and semi-supervised pretraining (CCT) all yield statistically equivalent performance for mean Intersection over Union (mIoU). We achieve a new state-of-the-art for SCDD and demonstrate that certain pretraining schemes result in superior performance on underrepresented classes. Additionally, we provide a large-scale unlabelled EL image dataset of 22000 images, and a 642-image labelled semantic segmentation EL dataset, for further research in developing self- and semi-supervised training techniques in this domain.

link: http://arxiv.org/abs/2402.17611v1

Neural Automated Writing Evaluation with Corrective Feedback

Izia Xiaoxiao Wang, Xihan Wu, Edith Coates, Min Zeng, Jiexin Kuang, Siliang Liu, Mengyang Qiu, Jungyeul Park

The utilization of technology in second language learning and teaching has become ubiquitous. For the assessment of writing specifically, automated writing evaluation (AWE) and grammatical error correction (GEC) have become immensely popular and effective methods for enhancing writing proficiency and delivering instant and individualized feedback to learners. By leveraging the power

of natural language processing (NLP) and machine learning algorithms, AWE and GEC systems have been developed separately to provide language learners with automated corrective feedback and more accurate and unbiased scoring that would otherwise be subject to examiners. In this paper, we propose an integrated system for automated writing evaluation with corrective feedback as a means of bridging the gap between AWE and GEC results for second language learners. This system enables language learners to simulate the essay writing tests: a student writes and submits an essay, and the system returns the assessment of the writing along with suggested grammatical error corrections. Given that automated scoring and grammatical correction are more efficient and cost-effective than human grading, this integrated system would also alleviate the burden of manually correcting innumerable essays.

link: http://arxiv.org/abs/2402.17613v1

Adapt Before Comparison: A New Perspective on Cross-Domain Few-Shot Segmentation

Jonas Herzog

Few-shot segmentation performance declines substantially when facing images from a domain different than the training domain, effectively limiting real-world use cases. To alleviate this, recently cross-domain few-shot segmentation (CD-FSS) has emerged. Works that address this task mainly attempted to learn segmentation on a source domain in a manner that generalizes across domains. Surprisingly, we can outperform these approaches while eliminating the training stage and removing their main segmentation network. We show test-time task-adaption is the key for successful CD-FSS instead. Task-adaption is achieved by appending small networks to the feature pyramid of a conventionally classification-pretrained backbone. To avoid overfitting to the few labeled samples in supervised fine-tuning, consistency across augmented views of input images serves as guidance while learning the parameters of the attached layers. Despite our self-restriction not to use any images other than the few labeled samples at test time, we achieve new state-of-the-art performance in CD-FSS, evidencing the need to rethink approaches for the task.

link: http://arxiv.org/abs/2402.17614v1