Fri 2024.05.03

Seeing Through the Clouds: Cloud Gap Imputation with Prithvi Foundation Model

Denys Godwin, Hanxi Li, Michael Cecil, Hamed Alemohammad

Filling cloudy pixels in multispectral satellite imagery is essential for accurate data analysis and downstream applications, especially for tasks which require time series data. To address this issue, we compare the performance of a foundational Vision Transformer (ViT) model with a baseline Conditional Generative Adversarial Network (CGAN) model for missing value imputation in time series of multispectral satellite imagery. We randomly mask time series of satellite images using real-world cloud masks and train each model to reconstruct the missing pixels. The ViT model is fine-tuned from a pretrained model, while the CGAN is trained from scratch. Using quantitative evaluation metrics such as structural similarity index and mean absolute error as well as qualitative visual analysis, we assess imputation accuracy and contextual preservation.

link: http://arxiv.org/abs/2404.19609v1

Radio Resource Management Design for RSMA: Optimization of Beamforming, User Admission, and Discrete/Continuous Rates with Imperfect SIC

L. F. Abanto-Leon, A. Krishnamoorthy, A. Garcia-Saavedra, G. H. Sim, R. Schober, M. Hollick

This paper investigates the radio resource management (RRM) design for multiuser rate-splitting multiple access (RSMA), accounting for various characteristics of practical wireless systems, such as the use of discrete rates, the inability to serve all users, and the imperfect successive interference cancellation (SIC). Specifically, failure to consider these characteristics in RRM design may lead to inefficient use of radio resources. Therefore, we formulate the RRM of RSMA as optimization problems to maximize respectively the weighted sum rate (WSR) and weighted energy efficiency (WEE), and jointly optimize the beamforming, user admission, discrete/continuous rates, accounting for imperfect SIC, which result in nonconvex mixed-integer nonlinear programs that are challenging to solve. Despite the difficulty of the optimization problems, we develop algorithms that can find high-quality solutions. We show via simulations that carefully accounting for the aforementioned characteristics, can lead to significant gains. Precisely, by considering that transmission rates are discrete, the transmit power can be utilized more intelligently, allocating just enough power to guarantee a given discrete rate. Additionally, we reveal that user admission plays a crucial role in RSMA, enabling additional gains compared to random admission by facilitating the servicing of selected users with mutually beneficial channel characteristics. Furthermore, provisioning for possibly imperfect SIC makes RSMA more robust and reliable.

link: http://arxiv.org/abs/2404.19611v1

Quantum Cloud Computing: Trends and Challenges

Muhammed Golec, Emir Sahin Hatay, Mustafa Golec, Murat Uyar, Merve Golec, Sukhpal Singh Gill

Quantum computing (QC) is a new paradigm that will revolutionize various areas of computing, especially cloud computing. QC, still in its infancy, is a costly technology capable of operating in highly isolated environments due to its rapid response to environmental factors. For this reason, it is still a challenging technology for researchers to reach. Integrating QC into an isolated remote server, like a cloud, and making it available to users can overcome these problems. Furthermore, experts predict that QC, with its ability to swiftly resolve complex and computationally intensive operations, will offer significant benefits in systems that process large amounts of data, like cloud computing. This article presents the vision and challenges for the quantum cloud computing (QCC) paradigm that will emerge with the integration of quantum and cloud computing. Next, we present the advantages of QC over classical computing applications. We analyze the effects of QC on cloud systems, such as cost, security, and scalability. Besides all of these advantages, we highlight research gaps in QCC, such as qubit stability and efficient resource allocation. This article identifies QCC's advantages and challenges for future research, highlighting research gaps.

link: http://arxiv.org/abs/2404.19612v1

SemiPL: A Semi-supervised Method for Event Sound Source Localization

Yue Li, Baiqiao Yin, Jinfu Liu, Jiajun Wen, Jiaying Lin, Mengyuan Liu

In recent years, Event Sound Source Localization has been widely applied in various fields. Recent works typically relying on the contrastive learning framework show impressive performance. However, all work is based on large relatively simple datasets. It's also crucial to understand and analyze human behaviors (actions and interactions of people), voices, and sounds in chaotic events in many applications, e.g., crowd management, and emergency response services. In this paper, we apply the existing model to a more complex dataset, explore the influence of parameters on the model, and propose a semi-supervised improvement method SemiPL. With the increase in data quantity and the influence of label quality, self-supervised learning will be an unstoppable trend. The experiment shows that the parameter adjustment will positively affect the existing model. In particular, SSPL achieved an improvement of 12.2% cIoU and 0.56% AUC in Chaotic World compared to the results provided. The code is available at: https://github.com/ly245422/SSPL

link: http://arxiv.org/abs/2404.19615v1

Be Aware of the Neighborhood Effect: Modeling Selection Bias under Interference

Haoxuan Li, Chunyuan Zheng, Sihao Ding, Peng Wu, Zhi Geng, Fuli Feng, Xiangnan He

Selection bias in recommender system arises from the recommendation process of system filtering and the interactive process of user selection. Many previous studies have focused on addressing selection bias to achieve unbiased learning of the prediction model, but ignore the fact that potential outcomes for a given user-item pair may vary with the treatments assigned to other user-item pairs, named neighborhood effect. To fill the gap, this paper formally formulates the neighborhood effect as an interference problem from the perspective of causal inference and introduces a treatment representation to capture the neighborhood effect. On this basis, we propose a novel ideal loss that can be used to deal with selection bias in the presence of neighborhood effect. We further develop two new estimators for estimating the proposed ideal loss. We theoretically establish the connection between the proposed and previous debiasing methods ignoring the neighborhood effect, showing that the proposed methods can achieve unbiased learning when both selection bias and neighborhood effect are present, while the existing methods are biased. Extensive semi-synthetic and real-world experiments are conducted to demonstrate the effectiveness of the proposed methods.

link: http://arxiv.org/abs/2404.19620v1

Acceso abierto en Argentina: una propuesta para el monitoreo de las publicaciones científicas con OpenAlex

Carolina Unzurrunzaga, Carolina Monti, Gastón Zalba, Juan Pablo Alperin

This study proposes a methodology using OpenAlex (OA) for tracking Open Access publications in the case of Argentina, a country where a self-archiving mandate has been in effect since 2013 (Law 26.899, 2013). A sample of 167,240 papers by researchers from the National Council for Scientific and Technical Research (CONICET) was created and analyzed using statistical techniques. We estimate that OA is able to capture between 85-93% of authors for all disciplines, with the exception of Social Sciences and Humanities, where it only reaches an estimated 47%. The availability of papers in Open Access was calculated to be 41% for the period 1953-2021 and 46% when considering exclusively the post-law period (2014-2021). In both periods, gold Open Access made up the most common route. When comparing equal periods post and pre-law, we observed that the upward trend of gold Open Access was pre-existing to the legislation and the availability of closed articles in repositories increased by 5% to what is estimated based on existing trends. However, while the green route has had a positive evolution, it has been the publication in gold journals that has boosted access to Argentine production more rapidly. We concluded that the OA-based methodology, piloted here for the first time, is viable for tracking Open Access in Argentina since it yields percentages similar to other national and international studies. En este

estudio se propone una metodolog\'ia utilizando OpenAlex (OA) para monitorear el acceso abierto (AA) a las publicaciones cient\'ificas para el caso de Argentina, pa\'is donde rige el mandato de autoarchivo -Ley 26.899 (2013)-. Se conform\'o una muestra con 167.240 art\'iculos de investigadores del Consejo Nacional de Investigaciones Cient\'ificas y T\'ecnicas (CONICET) que se analizaron con t\'ecnicas estad\'isticas. Se estim\'o que OA puede representar entre 85-93% de los autores para todas las disciplinas, excepto Ciencias Sociales y Humanidades, donde solo alcanza al 47%. Se calcul\'o que 41% de los art\'iculos publicados entre 1953-2021 incluidos en la fuente est\'an en AA, porcentaje que sube a 46% al considerar exclusivamente el periodo post ley (2014-2021). En ambos periodos es la v\'ia dorada la que representa mayor proporci\'on. Al comparar periodos iguales post y pre ley, se observ\'o que la tendencia en alza de la v\'ia dorada era preexistente a la legislaci\'on y la disponibilidad de art\'iculos cerrados en repositorios aument\'o un 5% a lo que se estima en base a tendencias existentes. Se concluye que si bien la v\'ia verde ha tenido una evoluci\'on positiva, ha sido la publicaci\'on en revistas doradas lo que ha impulsado m\'as r\'apidamente el acceso a la producci\'on argentina. Asimismo, que la metodolog\'ia basada en OA, piloteada aqu\'i por primera vez, es viable para monitorear el AA en Argentina ya que arroja porcentajes similares a otros estudios nacionales e internacionales.

link: http://arxiv.org/abs/2404.19627v1

Analyzing and Exploring Training Recipes for Large-Scale Transformer-Based Weather Prediction

Jared D. Willard, Peter Harrington, Shashank Subramanian, Ankur Mahesh, Travis A. O'Brien, William D. Collins

The rapid rise of deep learning (DL) in numerical weather prediction (NWP) has led to a proliferation of models which forecast atmospheric variables with comparable or superior skill than traditional physics-based NWP. However, among these leading DL models, there is a wide variance in both the training settings and architecture used. Further, the lack of thorough ablation studies makes it hard to discern which components are most critical to success. In this work, we show that it is possible to attain high forecast skill even with relatively off-the-shelf architectures, simple training procedures, and moderate compute budgets. Specifically, we train a minimally modified SwinV2 transformer on ERA5 data, and find that it attains superior forecast skill when compared against IFS. We present some ablations on key aspects of the training pipeline, exploring different loss functions, model sizes and depths, and multi-step fine-tuning to investigate their effect. We also examine the model performance with metrics beyond the typical ACC and RMSE, and investigate how the performance scales with model size.

link: http://arxiv.org/abs/2404.19630v1

On Training a Neural Network to Explain Binaries

Alexander Interrante-Grant, Andy Davis, Heather Preslier, Tim Leek

In this work, we begin to investigate the possibility of training a deep neural network on the task of binary code understanding. Specifically, the network would take, as input, features derived directly from binaries and output English descriptions of functionality to aid a reverse engineer in investigating the capabilities of a piece of closed-source software, be it malicious or benign. Given recent success in applying large language models (generative AI) to the task of source code summarization, this seems a promising direction. However, in our initial survey of the available datasets, we found nothing of sufficiently high quality and volume to train these complex models. Instead, we build our own dataset derived from a capture of Stack Overflow containing 1.1M entries. A major result of our work is a novel dataset evaluation method using the correlation between two distances on sample pairs: one distance in the embedding space of inputs and the other in the embedding space of outputs. Intuitively, if two samples have inputs close in the input embedding space, their outputs should also be close in the output embedding space. We found this Embedding Distance Correlation (EDC) test to be highly diagnostic, indicating that our collected dataset and several existing open-source datasets are of low quality as the distances are not well correlated. We proceed to explore the general applicability of EDC, applying it to a number of qualitatively known good datasets and a number of synthetically known bad ones and found it to be

a reliable indicator of dataset value.

link: http://arxiv.org/abs/2404.19631v1

DF Louvain: Fast Incrementally Expanding Approach for Community Detection on Dynamic Graphs

Subhajit Sahu

Community detection is the problem of recognizing natural divisions in networks. A relevant challenge in this problem is to find communities on rapidly evolving graphs. In this report we present our Parallel Dynamic Frontier (DF) Louvain algorithm, which given a batch update of edge deletions and insertions, incrementally identifies and processes an approximate set of affected vertices in the graph with minimal overhead, while using a novel approach of incrementally updating weighted-degrees of vertices and total edge weights of communities. We also present our parallel implementations of Naive-dynamic (ND) and Delta-screening (DS) Louvain. On a server with a 64-core AMD EPYC-7742 processor, our experiments show that DF Louvain obtains speedups of 179x, 7.2x, and 5.3x on real-world dynamic graphs, compared to Static, ND, and DS Louvain, respectively, and is 183x, 13.8x, and 8.7x faster, respectively, on large graphs with random batch updates. Moreover, DF Louvain improves its performance by 1.6x for every doubling of threads.

link: http://arxiv.org/abs/2404.19634v1

SpComm3D: A Framework for Enabling Sparse Communication in 3D Sparse Kernels

Nabil Abubaker, Torsten Hoefler

Existing 3D algorithms for distributed-memory sparse kernels suffer from limited scalability due to reliance on bulk sparsity-agnostic communication. While easier to use, sparsity-agnostic communication leads to unnecessary bandwidth and memory consumption. We present SpComm3D, a framework for enabling sparsity-aware communication and minimal memory footprint such that no unnecessary data is communicated or stored in memory. SpComm3D performs sparse communication efficiently with minimal or no communication buffers to further reduce memory consumption. SpComm3D detaches the local computation at each processor from the communication, allowing flexibility in choosing the best accelerated version for computation. We build 3D algorithms with SpComm3D for the two important sparse ML kernels: Sampled Dense-Dense Matrix Multiplication (SDDMM) and Sparse matrix-matrix multiplication (SpMM). Experimental evaluations on up to 1800 processors demonstrate that SpComm3D has superior scalability and outperforms state-of-the-art sparsity-agnostic methods with up to 20x improvement in terms of communication, memory, and runtime of SDDMM and SpMM. The code is available at: https://github.com/nfabubaker/SpComm3D

link: http://arxiv.org/abs/2404.19638v1

ESP-Zero: Unsupervised enhancement of zero-shot classification for Extremely Sparse Point cloud

Jiayi Han, Zidi Cao, Weibo Zheng, Xiangguo Zhou, Xiangjian He, Yuanfang Zhang, Daisen Wei

In recent years, zero-shot learning has attracted the focus of many researchers, due to its flexibility and generality. Many approaches have been proposed to achieve the zero-shot classification of the point clouds for 3D object understanding, following the schema of CLIP. However, in the real world, the point clouds could be extremely sparse, dramatically limiting the effectiveness of the 3D point cloud encoders, and resulting in the misalignment of point cloud features and text embeddings. To the point cloud encoders to fit the extremely sparse point clouds without re-running the pre-training procedure which could be time-consuming and expensive, in this work, we propose an unsupervised model adaptation approach to enhance the point cloud encoder for the extremely sparse point clouds. We propose a novel fused-cross attention layer that expands the pre-trained self-attention layer with additional learnable tokens and attention blocks, which effectively modifies the point cloud features while maintaining the alignment between point cloud features and text embeddings. We also propose a complementary learning-based self-distillation schema that

encourages the modified features to be pulled apart from the irrelevant text embeddings without overfitting the feature space to the observed text embeddings. Extensive experiments demonstrate that the proposed approach effectively increases the zero-shot capability on extremely sparse point clouds, and overwhelms other state-of-the-art model adaptation approaches.

link: http://arxiv.org/abs/2404.19639v1

MetaCoCo: A New Few-Shot Classification Benchmark with Spurious Correlation

Min Zhang, Haoxuan Li, Fei Wu, Kun Kuang

Out-of-distribution (OOD) problems in few-shot classification (FSC) occur when novel classes sampled from testing distributions differ from base classes drawn from training distributions, which considerably degrades the performance of deep learning models deployed in real-world applications. Recent studies suggest that the OOD problems in FSC mainly including: (a) cross-domain few-shot classification (CD-FSC) and (b) spurious-correlation few-shot classification (SC-FSC). Specifically, CD-FSC occurs when a classifier learns transferring knowledge from base classes drawn from seen training distributions but recognizes novel classes sampled from unseen testing distributions. In contrast, SC-FSC arises when a classifier relies on non-causal features (or contexts) that happen to be correlated with the labels (or concepts) in base classes but such relationships no longer hold during the model deployment. Despite CD-FSC has been extensively studied, SC-FSC remains understudied due to lack of the corresponding evaluation benchmarks. To this end, we present Meta Concept Context (MetaCoCo), a benchmark with spurious-correlation shifts collected from real-world scenarios. Moreover, to quantify the extent of spurious-correlation shifts of the presented MetaCoCo, we further propose a metric by using CLIP as a pre-trained vision-language model. Extensive experiments on the proposed benchmark are performed to evaluate the state-of-the-art methods in FSC, cross-domain shifts, and self-supervised learning. The experimental results show that the performance of the existing methods degrades significantly in the presence of spurious-correlation shifts. We open-source all codes of our benchmark and hope that the proposed MetaCoCo can facilitate future research on spurious-correlation shifts problems in FSC. The code is available at: https://github.com/remiMZ/MetaCoCo-ICLR24.

link: http://arxiv.org/abs/2404.19644v1

Provably Robust Conformal Prediction with Improved Efficiency

Ge Yan, Yaniv Romano, Tsui-Wei Weng

Conformal prediction is a powerful tool to generate uncertainty sets with guaranteed coverage using any predictive model, under the assumption that the training and test data are i.i.d.. Recently, it has been shown that adversarial examples are able to manipulate conformal methods to construct prediction sets with invalid coverage rates, as the i.i.d. assumption is violated. To address this issue, a recent work, Randomized Smoothed Conformal Prediction (RSCP), was first proposed to certify the robustness of conformal prediction methods to adversarial noise. However, RSCP has two major limitations: (i) its robustness guarantee is flawed when used in practice and (ii) it tends to produce large uncertainty sets. To address these limitations, we first propose a novel framework called RSCP+ to provide provable robustness guarantee in evaluation, which fixes the issues in the original RSCP method. Next, we propose two novel methods, Post-Training Transformation (PTT) and Robust Conformal Training (RCT), to effectively reduce prediction set size with little computation overhead. Experimental results in CIFAR10, CIFAR100, and ImageNet suggest the baseline method only yields trivial predictions including full label set, while our methods could boost the efficiency by up to $4.36\times$, $5.46\times$, and $16.9\times$ respectively and provide practical robustness guarantee. Our codes are available at https://github.com/Trustworthy-ML-Lab/Provably-Robust-Conformal-Prediction.

link: http://arxiv.org/abs/2404.19651v1

VimTS: A Unified Video and Image Text Spotter for Enhancing the Cross-domain Generalization

Yuliang Liu, Mingxin Huang, Hao Yan, Linger Deng, Weijia Wu, Hao Lu, Chunhua Shen, Lianwen Jin, Xiang Bai

Text spotting, a task involving the extraction of textual information from image or video sequences, faces challenges in cross-domain adaption, such as image-to-image and image-to-video generalization. In this paper, we introduce a new method, termed VimTS, which enhances the generalization ability of the model by achieving better synergy among different tasks. Typically, we propose a Prompt Queries Generation Module and a Tasks-aware Adapter to effectively convert the original single-task model into a multi-task model suitable for both image and video scenarios with minimal additional parameters. The Prompt Queries Generation Module facilitates explicit interaction between different tasks, while the Tasks-aware Adapter helps the model dynamically learn suitable features for each task. Additionally, to further enable the model to learn temporal information at a lower cost, we propose a synthetic video text dataset (VTD-368k) by leveraging the Content Deformation Fields (CoDeF) algorithm. Notably, our method outperforms the state-of-the-art method by an average of 2.6% in six cross-domain benchmarks such as TT-to-IC15, CTW1500-to-TT, and TT-to-CTW1500. For video-level cross-domain adaption, our method even surpasses the previous end-to-end video spotting method in ICDAR2015 video and DSText v2 by an average of 5.5% on the MOTA metric, using only image-level data. We further demonstrate that existing Large Multimodal Models exhibit limitations in generating cross-domain scene text spotting, in contrast to our VimTS model which requires significantly fewer parameters and data. The code and datasets will be made available at the https://VimTextSpotter.github.io.

link: http://arxiv.org/abs/2404.19652v1

Masked Multi-Query Slot Attention for Unsupervised Object Discovery

Rishav Pramanik, José-Fabian Villa-Vásquez, Marco Pedersoli

Unsupervised object discovery is becoming an essential line of research for tackling recognition problems that require decomposing an image into entities, such as semantic segmentation and object detection. Recently, object-centric methods that leverage self-supervision have gained popularity, due to their simplicity and adaptability to different settings and conditions. However, those methods do not exploit effective techniques already employed in modern self-supervised approaches. In this work, we consider an object-centric approach in which DINO ViT features are reconstructed via a set of queried representations called slots. Based on that, we propose a masking scheme on input features that selectively disregards the background regions, inducing our model to focus more on salient objects during the reconstruction phase. Moreover, we extend the slot attention to a multi-query approach, allowing the model to learn multiple sets of slots, producing more stable masks. During training, these multiple sets of slots are learned independently while, at test time, these sets are merged through Hungarian matching to obtain the final slots. Our experimental results and ablations on the PASCAL-VOC 2012 dataset show the importance of each component and highlight how their combination consistently improves object localization. Our source code is available at: https://github.com/rishavpramanik/maskedmultiqueryslot

link: http://arxiv.org/abs/2404.19654v1