Wed 2024.05.01

4DBInfer: A 4D Benchmarking Toolbox for Graph-Centric Predictive Modeling on Relational DBs

Minjie Wang, Quan Gan, David Wipf, Zhenkun Cai, Ning Li, Jianheng Tang, Yanlin Zhang, Zizhao Zhang, Zunyao Mao, Yakun Song, Yanbo Wang, Jiahang Li, Han Zhang, Guang Yang, Xiao Qin, Chuan Lei, Muhan Zhang, Weinan Zhang, Christos Faloutsos, Zheng Zhang

Although RDBs store vast amounts of rich, informative data spread across interconnected tables, the progress of predictive machine learning models as applied to such tasks arguably falls well behind advances in other domains such as computer vision or natural language processing. This deficit stems, at least in part, from the lack of established/public RDB benchmarks as needed for training and evaluation purposes. As a result, related model development thus far often defaults to tabular approaches trained on ubiquitous single-table benchmarks, or on the relational side, graph-based alternatives such as GNNs applied to a completely different set of graph datasets devoid of tabular characteristics. To more precisely target RDBs lying at the nexus of these two complementary regimes, we explore a broad class of baseline models predicated on: (i) converting multi-table datasets into graphs using various strategies equipped with efficient subsampling, while preserving tabular characteristics; and (ii) trainable models with well-matched inductive biases that output predictions based on these input subgraphs. Then, to address the dearth of suitable public benchmarks and reduce siloed comparisons, we assemble a diverse collection of (i) large-scale RDB datasets and (ii) coincident predictive tasks. From a delivery standpoint, we operationalize the above four dimensions (4D) of exploration within a unified, scalable open-source toolbox called 4DBInfer. We conclude by presenting evaluations using 4DBInfer, the results of which highlight the importance of considering each such dimension in the design of RDB predictive models, as well as the limitations of more naive approaches such as simply joining adjacent tables. Our source code is released at https://github.com/awslabs/multi-table-benchmark .

link: http://arxiv.org/abs/2404.18209v1

A survey of dynamic graph neural networks

Yanping Zheng, Lu Yi, Zhewei Wei

Graph neural networks (GNNs) have emerged as a powerful tool for effectively mining and learning from graph-structured data, with applications spanning numerous domains. However, most research focuses on static graphs, neglecting the dynamic nature of real-world networks where topologies and attributes evolve over time. By integrating sequence modeling modules into traditional GNN architectures, dynamic GNNs aim to bridge this gap, capturing the inherent temporal dependencies of dynamic graphs for a more authentic depiction of complex networks. This paper provides a comprehensive review of the fundamental concepts, key techniques, and state-of-the-art dynamic GNN models. We present the mainstream dynamic GNN models in detail and categorize models based on how temporal information is incorporated. We also discuss large-scale dynamic GNNs and pre-training techniques. Although dynamic GNNs have shown superior performance, challenges remain in scalability, handling heterogeneous information, and lack of diverse graph datasets. The paper also discusses possible future directions, such as adaptive and memory-enhanced models, inductive learning, and theoretical analysis.

link: http://arxiv.org/abs/2404.18211v1

Paint by Inpaint: Learning to Add Image Objects by Removing Them First

Navve Wasserman, Noam Rotstein, Roy Ganz, Ron Kimmel

Image editing has advanced significantly with the introduction of text-conditioned diffusion models. Despite this progress, seamlessly adding objects to images based on textual instructions without requiring user-provided input masks remains a challenge. We address this by leveraging the insight that removing objects (Inpaint) is significantly simpler than its inverse process of adding them (Paint), attributed to the utilization of segmentation mask datasets alongside inpainting models that

inpaint within these masks. Capitalizing on this realization, by implementing an automated and extensive pipeline, we curate a filtered large-scale image dataset containing pairs of images and their corresponding object-removed versions. Using these pairs, we train a diffusion model to inverse the inpainting process, effectively adding objects into images. Unlike other editing datasets, ours features natural target images instead of synthetic ones; moreover, it maintains consistency between source and target by construction. Additionally, we utilize a large Vision-Language Model to provide detailed descriptions of the removed objects and a Large Language Model to convert these descriptions into diverse, natural-language instructions. We show that the trained model surpasses existing ones both qualitatively and quantitatively, and release the large-scale dataset alongside the trained models for the community.

link: http://arxiv.org/abs/2404.18212v1

S^{2}Mamba: A Spatial-spectral State Space Model for Hyperspectral Image Classification

Guanchun Wang, Xiangrong Zhang, Zelin Peng, Tianyang Zhang, Xiuping Jia, Licheng Jiao

Land cover analysis using hyperspectral images (HSI) remains an open problem due to their low spatial resolution and complex spectral information. Recent studies are primarily dedicated to designing Transformer-based architectures for spatial-spectral long-range dependencies modeling, which is computationally expensive with quadratic complexity. Selective structured state space model (Mamba), which is efficient for modeling long-range dependencies with linear complexity, has recently shown promising progress. However, its potential in hyperspectral image processing that requires handling numerous spectral bands has not yet been explored. In this paper, we innovatively propose S^{2}Mamba, a spatial-spectral state space model for hyperspectral image classification, to excavate spatial-spectral contextual features, resulting in more efficient and accurate land cover analysis. In S^{2}Mamba, two selective structured state space models through different dimensions are designed for feature extraction, one for spatial, and the other for spectral, along with a spatial-spectral mixture gate for optimal fusion. More specifically, S^{2}Mamba first captures spatial contextual relations by interacting each pixel with its adjacent through a Patch Cross Scanning module and then explores semantic information from continuous spectral bands through a Bi-directional Spectral Scanning module. Considering the distinct expertise of the two attributes in homogenous and complicated texture scenes, we realize the Spatial-spectral Mixture Gate by a group of learnable matrices, allowing for the adaptive incorporation of representations learned across different dimensions. Extensive experiments conducted on HSI classification benchmarks demonstrate the superiority and prospect of S^{2}Mamba. The code will be available at: https://github.com/PURE-melo/S2Mamba.

link: http://arxiv.org/abs/2404.18213v1

Contrastive Learning Method for Sequential Recommendation based on Multi-Intention Disentanglement

Zeyu Hu, Yuzhi Xiao, Tao Huang, Xuanrong Huo

Sequential recommendation is one of the important branches of recommender system, aiming to achieve personalized recommended items for the future through the analysis and prediction of users' ordered historical interactive behaviors. However, along with the growth of the user volume and the increasingly rich behavioral information, how to understand and disentangle the user's interactive multi-intention effectively also poses challenges to behavior prediction and sequential recommendation. In light of these challenges, we propose a Contrastive Learning sequential recommendation method based on Multi-Intention Disentanglement (MIDCL). In our work, intentions are recognized as dynamic and diverse, and user behaviors are often driven by current multi-intentions, which means that the model needs to not only mine the most relevant implicit intention for each user, but also impair the influence from irrelevant intentions. Therefore, we choose Variational Auto-Encoder (VAE) to realize the disentanglement of users' multi-intentions, and propose two types of contrastive learning paradigms for finding the most relevant user's interactive intention, and maximizing the mutual information of positive sample pairs, respectively. Experimental results show that MIDCL not only has significant superiority over most existing

baseline methods, but also brings a more interpretable case to the research about intention-based prediction and recommendation.

link: http://arxiv.org/abs/2404.18214v1

L3Cube-MahaNews: News-based Short Text and Long Document Classification Datasets in Marathi

Saloni Mittal, Vidula Magdum, Omkar Dhekane, Sharayu Hiwarkhedkar, Raviraj Joshi

The availability of text or topic classification datasets in the low-resource Marathi language is limited, typically consisting of fewer than 4 target labels, with some achieving nearly perfect accuracy. In this work, we introduce L3Cube-MahaNews, a Marathi text classification corpus that focuses on News headlines and articles. This corpus stands out as the largest supervised Marathi Corpus, containing over 1.05L records classified into a diverse range of 12 categories. To accommodate different document lengths, MahaNews comprises three supervised datasets specifically designed for short text, long documents, and medium paragraphs. The consistent labeling across these datasets facilitates document length-based analysis. We provide detailed data statistics and baseline results on these datasets using state-of-the-art pre-trained BERT models. We conduct a comparative analysis between monolingual and multilingual BERT models, including MahaBERT, IndicBERT, and MuRIL. The monolingual MahaBERT model outperforms all others on every dataset. These resources also serve as Marathi topic classification datasets or models and are publicly available at https://github.com/l3cube-pune/MarathiNLP .

link: http://dx.doi.org/10.1007/978-3-031-58495-4_4

A quantum compiler design method by using linear combinations of permutations

Ammar Daskin

A matrix can be converted into a doubly stochastic matrix by using two diagonal matrices. And a doubly stochastic matrix can be written as a sum of permutation matrices. In this paper, we describe a method to write a given generic matrix in terms of quantum gates based on the block encoding. In particular, we first show how to convert a matrix into doubly stochastic matrices and by using Birkhoff's algorithm, we express that matrix in terms of a linear combination of permutations which can be mapped to quantum circuits. We then discuss a few optimization techniques that can be applied in a possibly future quantum compiler software based on the method described here.

link: http://arxiv.org/abs/2404.18226v1

TextGram: Towards a better domain-adaptive pretraining

Sharayu Hiwarkhedkar, Saloni Mittal, Vidula Magdum, Omkar Dhekane, Raviraj Joshi, Geetanjali Kale, Arnav Ladkat

For green AI, it is crucial to measure and reduce the carbon footprint emitted during the training of large language models. In NLP, performing pre-training on Transformer models requires significant computational resources. This pre-training involves using a large amount of text data to gain prior knowledge for performing downstream tasks. Thus, it is important that we select the correct data in the form of domain-specific data from this vast corpus to achieve optimum results aligned with our domain-specific tasks. While training on large unsupervised data is expensive, it can be optimized by performing a data selection step before pretraining. Selecting important data reduces the space overhead and the substantial amount of time required to pre-train the model while maintaining constant accuracy. We investigate the existing selection strategies and propose our own domain-adaptive data selection method - TextGram - that effectively selects essential data from large corpora. We compare and evaluate the results of finetuned models for text classification task with and without data selection. We show that the proposed strategy works better compared to other selection methods.

link: http://dx.doi.org/10.1007/978-3-031-58495-4_12

From Persona to Personalization: A Survey on Role-Playing Language Agents

Jiangjie Chen, Xintao Wang, Rui Xu, Siyu Yuan, Yikai Zhang, Wei Shi, Jian Xie, Shuang Li, Ruihan Yang, Tinghui Zhu, Aili Chen, Nianqi Li, Lida Chen, Caiyu Hu, Siye Wu, Scott Ren, Ziquan Fu, Yanghua Xiao

Recent advancements in large language models (LLMs) have significantly boosted the rise of Role-Playing Language Agents (RPLAs), i.e., specialized AI systems designed to simulate assigned personas. By harnessing multiple advanced abilities of LLMs, including in-context learning, instruction following, and social intelligence, RPLAs achieve a remarkable sense of human likeness and vivid role-playing performance. RPLAs can mimic a wide range of personas, ranging from historical figures and fictional characters to real-life individuals. Consequently, they have catalyzed numerous AI applications, such as emotional companions, interactive video games, personalized assistants and copilots, and digital clones. In this paper, we conduct a comprehensive survey of this field, illustrating the evolution and recent progress in RPLAs integrating with cutting-edge LLM technologies. We categorize personas into three types: 1) Demographic Persona, which leverages statistical stereotypes; 2) Character Persona, focused on well-established figures; and 3) Individualized Persona, customized through ongoing user interactions for personalized services. We begin by presenting a comprehensive overview of current methodologies for RPLAs, followed by the details for each persona type, covering corresponding data sourcing, agent construction, and evaluation. Afterward, we discuss the fundamental risks, existing limitations, and future prospects of RPLAs. Additionally, we provide a brief review of RPLAs in AI applications, which reflects practical user demands that shape and drive RPLA research. Through this work, we aim to establish a clear taxonomy of RPLA research and applications, and facilitate future research in this critical and ever-evolving field, and pave the way for a future where humans and RPLAs coexist in harmony.

link: http://arxiv.org/abs/2404.18231v1

A Note on Asynchronous Challenges: Unveiling Formulaic Bias and Data Loss in the Hayashi-Yoshida Estimator

Evangelos Georgiadis

The Hayashi-Yoshida (\HY)-estimator exhibits an intrinsic, telescoping property that leads to an often overlooked computational bias, which we denote,formulaic or intrinsic bias. This formulaic bias results in data loss by cancelling out potentially relevant data points, the nonextant data points. This paper attempts to formalize and quantify the data loss arising from this bias. In particular, we highlight the existence of nonextant data points via a concrete example, and prove necessary and sufficient conditions for the telescoping property to induce this type of formulaic bias.Since this type of bias is nonexistent when inputs, i.e., observation times, $\Pi^{(1)} :=(t_i^{(1)})_{i=0,1,\ldots}$ and $\Pi^{(2)} :=(t_j^{(2)})_{j=0,1,\ldots}$, are synchronous, we introduce the (a,b)-asynchronous adversary. This adversary generates inputs $\Pi^{(1)}$ and $\Pi^{(2)}$ according to two independent homogenous Poisson processes with rates a>0 and b>0, respectively. We address the foundational questions regarding cumulative minimal (or least) average data point loss, and determine the values for a and b. We prove that for equal rates a=b, the minimal average cumulative data loss over both inputs is attained and amounts to 25\%. We present an algorithm, which is based on our theorem, for computing the exact number of nonextant data points given inputs $\Pi^{(1)}$ and $\Pi^{(2)}$, and suggest alternative methods. Finally, we use simulated data to empirically compare the (cumulative) average data loss of the (\HY)-estimator.

link: http://arxiv.org/abs/2404.18233v1

Flood Data Analysis on SpaceNet 8 Using Apache Sedona

Yanbing Bai, Zihao Yang, Jinze Yu, Rui-Yang Ju, Bin Yang, Erick Mas, Shunichi Koshimura

With the escalating frequency of floods posing persistent threats to human life and property, satellite remote sensing has emerged as an indispensable tool for monitoring flood hazards. SpaceNet8 offers a unique opportunity to leverage cutting-edge artificial intelligence technologies to assess these hazards. A significant contribution of this research is its application of Apache Sedona, an advanced platform specifically designed for the efficient and distributed processing of large-scale geospatial data. This platform aims to enhance the efficiency of error analysis, a critical

aspect of improving flood damage detection accuracy. Based on Apache Sedona, we introduce a novel approach that addresses the challenges associated with inaccuracies in flood damage detection. This approach involves the retrieval of cases from historical flood events, the adaptation of these cases to current scenarios, and the revision of the model based on clustering algorithms to refine its performance. Through the replication of both the SpaceNet8 baseline and its top-performing models, we embark on a comprehensive error analysis. This analysis reveals several main sources of inaccuracies. To address these issues, we employ data visual interpretation and histogram equalization techniques, resulting in significant improvements in model metrics. After these enhancements, our indicators show a notable improvement, with precision up by 5%, F1 score by 2.6%, and IoU by 4.5%. This work highlights the importance of advanced geospatial data processing tools, such as Apache Sedona. By improving the accuracy and efficiency of flood detection, this research contributes to safeguarding public safety and strengthening infrastructure resilience in flood-prone areas, making it a valuable addition to the field of remote sensing and disaster management.

link: http://arxiv.org/abs/2404.18235v1

SOUL: Unlocking the Power of Second-Order Optimization for LLM Unlearning

Jinghan Jia, Yihua Zhang, Yimeng Zhang, Jiancheng Liu, Bharat Runwal, James Diffenderfer, Bhavya Kailkhura, Sijia Liu

Large Language Models (LLMs) have highlighted the necessity of effective unlearning mechanisms to comply with data regulations and ethical AI practices. LLM unlearning aims at removing undesired data influences and associated model capabilities without compromising utility out of the scope of unlearning. While interest in studying LLM unlearning is growing,the impact of the optimizer choice for LLM unlearning remains under-explored. In this work, we shed light on the significance of optimizer selection in LLM unlearning for the first time, establishing a clear connection between {second-order optimization} and influence unlearning (a classical approach using influence functions to update the model for data influence removal). This insight propels us to develop a second-order unlearning framework, termed SOUL, built upon the second-order clipped stochastic optimization (Sophia)-based LLM training method. SOUL extends the static, one-shot model update using influence unlearning to a dynamic, iterative unlearning process. Our extensive experiments show that SOUL consistently outperforms conventional first-order methods across various unlearning tasks, models, and metrics, suggesting the promise of second-order optimization in providing a scalable and easily implementable solution for LLM unlearning.

link: http://arxiv.org/abs/2404.18239v1

LEGENT: Open Platform for Embodied Agents

Zhili Cheng, Zhitong Wang, Jinyi Hu, Shengding Hu, An Liu, Yuge Tu, Pengkai Li, Lei Shi, Zhiyuan Liu, Maosong Sun

Despite advancements in Large Language Models (LLMs) and Large Multimodal Models (LMMs), their integration into language-grounded, human-like embodied agents remains incomplete, hindering complex real-life task performance in physical environments. Existing integrations often feature limited open sourcing, challenging collective progress in this field. We introduce LEGENT, an open, scalable platform for developing embodied agents using LLMs and LMMs. LEGENT offers a dual approach: a rich, interactive 3D environment with communicable and actionable agents, paired with a user-friendly interface, and a sophisticated data generation pipeline utilizing advanced algorithms to exploit supervision from simulated worlds at scale. In our experiments, an embryonic vision-language-action model trained on LEGENT-generated data surpasses GPT-4V in embodied tasks, showcasing promising generalization capabilities.

link: http://arxiv.org/abs/2404.18243v1

FAD-SAR: A Novel Fishing Activity Detection System via Synthetic Aperture Radar Images Based on Deep Learning Method

Yanbing Bai, Rui-Yang Ju, Siao Li, Zihao Yang, Jinze Yu

Illegal, unreported, and unregulated (IUU) fishing seriously affects various aspects of human life. However, current methods for detecting and monitoring IUU activities at sea have limitations. While Synthetic Aperture Radar (SAR) can complement existing vessel detection systems, extracting useful information from SAR images using traditional methods, especially for IUU fishing identification, poses challenges. This paper proposes a deep learning-based system for detecting fishing activities. We implemented this system on the xView3 dataset using six classical object detection models: Faster R-CNN, Cascade R-CNN, SSD, RetinaNet, FSAF, and FCOS. We applied improvement methods to enhance the performance of the Faster R-CNN model. Specifically, training the Faster R-CNN model using Online Hard Example Mining (OHEM) strategy improved the Avg-F1 value from 0.212 to 0.216, representing a 1.96% improvement.

link: http://arxiv.org/abs/2404.18245v1

AdaFSNet: Time Series Classification Based on Convolutional Network with a Adaptive and Effective Kernel Size Configuration

Haoxiao Wang, Bo Peng, Jianhua Zhang, Xu Cheng

Time series classification is one of the most critical and challenging problems in data mining, existing widely in various fields and holding significant research importance. Despite extensive research and notable achievements with successful real-world applications, addressing the challenge of capturing the appropriate receptive field (RF) size from one-dimensional or multi-dimensional time series of varying lengths remains a persistent issue, which greatly impacts performance and varies considerably across different datasets. In this paper, we propose an Adaptive and Effective Full-Scope Convolutional Neural Network (AdaFSNet) to enhance the accuracy of time series classification. This network includes two Dense Blocks. Particularly, it can dynamically choose a range of kernel sizes that effectively encompass the optimal RF size for various datasets by incorporating multiple prime numbers corresponding to the time series length. We also design a TargetDrop block, which can reduce redundancy while extracting a more effective RF. To assess the effectiveness of the AdaFSNet network, comprehensive experiments were conducted using the UCR and UEA datasets, which include one-dimensional and multi-dimensional time series data, respectively. Our model surpassed baseline models in terms of classification accuracy, underscoring the AdaFSNet network's efficiency and effectiveness in handling time series classification tasks.

link: http://arxiv.org/abs/2404.18246v1