Wed 2024.02.28

Cross-Modal Contextualized Diffusion Models for Text-Guided Visual Generation and Editing

Ling Yang, Zhilong Zhang, Zhaochen Yu, Jingwei Liu, Minkai Xu, Stefano Ermon, Bin Cui

Conditional diffusion models have exhibited superior performance in high-fidelity text-guided visual generation and editing. Nevertheless, prevailing text-guided visual diffusion models primarily focus on incorporating text-visual relationships exclusively into the reverse process, often disregarding their relevance in the forward process. This inconsistency between forward and reverse processes may limit the precise conveyance of textual semantics in visual synthesis results. To address this issue, we propose a novel and general contextualized diffusion model (ContextDiff) by incorporating the cross-modal context encompassing interactions and alignments between text condition and visual sample into forward and reverse processes. We propagate this context to all timesteps in the two processes to adapt their trajectories, thereby facilitating cross-modal conditional modeling. We generalize our contextualized diffusion to both DDPMs and DDIMs with theoretical derivations, and demonstrate the effectiveness of our model in evaluations with two challenging tasks: text-to-image generation, and text-to-video editing. In each task, our ContextDiff achieves new state-of-the-art performance, significantly enhancing the semantic alignment between text condition and generated samples, as evidenced by quantitative and qualitative evaluations. Our code is available at https://github.com/YangLing0818/ContextDiff

link: http://arxiv.org/abs/2402.16627v1

GenAINet: Enabling Wireless Collective Intelligence via Knowledge Transfer and Reasoning

Hang Zou, Qiyang Zhao, Lina Bariah, Yu Tian, Mehdi Bennis, Samson Lasaulce, Merouane Debbah, Faouzi Bader

Generative artificial intelligence (GenAI) and communication networks are expected to have groundbreaking synergies in 6G. Connecting GenAI agents over a wireless network can potentially unleash the power of collective intelligence and pave the way for artificial general intelligence (AGI). However, current wireless networks are designed as a "data pipe" and are not suited to accommodate and leverage the power of GenAI. In this paper, we propose the GenAINet framework in which distributed GenAI agents communicate knowledge (high-level concepts or abstracts) to accomplish arbitrary tasks. We first provide a network architecture integrating GenAI capabilities to manage both network protocols and applications. Building on this, we investigate effective communication and reasoning problems by proposing a semantic-native GenAINet. Specifically, GenAI agents extract semantic concepts from multi-modal raw data, build a knowledgebase representing their semantic relations, which is retrieved by GenAI models for planning and reasoning. Under this paradigm, an agent can learn fast from other agents' experience for making better decisions with efficient communications. Furthermore, we conduct two case studies where in wireless device query, we show that extracting and transferring knowledge can improve query accuracy with reduced communication; and in wireless power control, we show that distributed agents can improve decisions via collaborative reasoning. Finally, we address that developing a hierarchical semantic level Telecom world model is a key path towards network of collective intelligence.

link: http://arxiv.org/abs/2402.16631v1

Domain Embeddings for Generating Complex Descriptions of Concepts in Italian Language

Alessandro Maisto

In this work, we propose a Distributional Semantic resource enriched with linguistic and lexical information extracted from electronic dictionaries, designed to address the challenge of bridging the gap between the continuous semantic values represented by distributional vectors and the discrete descriptions offered by general semantics theory. Recently, many researchers have concentrated on the nexus between embeddings and a comprehensive theory of semantics and meaning. This often involves decoding the representation of word meanings in Distributional Models into a set of discrete, manually constructed properties such as semantic primitives or features, using neural decoding techniques. Our approach introduces an alternative strategy grounded in linguistic data. We have developed a collection of domain-specific co-occurrence matrices, derived from two sources: a classification of Italian nouns categorized into 4 semantic traits and 20 concrete noun sub-categories, and a list of Italian verbs classified according to their semantic classes. In these matrices, the co-occurrence values for each word are calculated exclusively with a defined set of words pertinent to a particular lexical domain. The resource comprises 21 domain-specific matrices, one comprehensive matrix, and a Graphical User Interface. Our model facilitates the generation of reasoned semantic descriptions of concepts by selecting matrices directly associated with concrete conceptual knowledge, such as a matrix based on location nouns and the concept of animal habitats. We assessed the utility of the resource through two experiments, achieving promising outcomes in both: the automatic classification of animal nouns and the extraction of animal features.

link: http://arxiv.org/abs/2402.16632v1

Differentiable Particle Filtering using Optimal Placement Resampling

Domonkos Csuzdi, Olivér Tör■, Tamás Bécsi

Particle filters are a frequent choice for inference tasks in nonlinear and non-Gaussian state-space models. They can either be used for state inference by approximating the filtering distribution or for parameter inference by approximating the marginal data (observation) likelihood. A good proposal distribution and a good resampling scheme are crucial to obtain low variance estimates. However, traditional methods like multinomial resampling introduce nondifferentiability in PF-based loss functions for parameter estimation, prohibiting gradient-based learning tasks. This work proposes a differentiable resampling scheme by deterministic sampling from an empirical cumulative distribution function. We evaluate our method on parameter inference tasks and proposal learning.

link: http://arxiv.org/abs/2402.16639v1

DRSI-Net: Dual-Residual Spatial Interaction Network for Multi-Person Pose Estimation

Shang Wu, Bin Wang

Multi-person pose estimation (MPPE), which aims to locate keypoints for all persons in the frames, is an active research branch of computer vision. Variable human poses and complex scenes make MPPE dependent on both local details and global structures, and the absence of them may cause keypoint feature misalignment. In this case, high-order spatial interactions that can effectively link the local and global information of features are particularly important. However, most methods do not have spatial interactions, and a few methods have low-order spatial interactions but they are difficult to achieve a good balance between accuracy and complexity. To address the above problems, a Dual-Residual Spatial Interaction Network (DRSI-Net) for MPPE with high accuracy and low complexity is proposed in this paper. DRSI-Net recursively performs residual spatial information interactions on neighbor features, so that more useful spatial information can be retained and more similarities can be obtained between shallow and deep extracted features. The channel and spatial dual attention mechanism introduced in the multi-scale feature fusion also helps the network to adaptively focus on features relevant to target keypoints and further refine generated poses. At the same time, by optimizing interactive channel dimensions and dividing gradient flow, the spatial interaction module is designed to be lightweight, which reduces the complexity of the network. According to the experimental results on the COCO dataset, the proposed DRSI-Net outperforms other state-of-the-art methods in both accuracy and complexity.

link: http://arxiv.org/abs/2402.16640v1

Towards Open-ended Visual Quality Comparison

Haoning Wu, Hanwei Zhu, Zicheng Zhang, Erli Zhang, Chaofeng Chen, Liang Liao, Chunyi Li, Annan Wang, Wenxiu Sun, Qiong Yan, Xiaohong Liu, Guangtao Zhai, Shiqi Wang, Weisi Lin

Comparative settings (e.g. pairwise choice, listwise ranking) have been adopted by a wide range of subjective studies for image quality assessment (IQA), as it inherently standardizes the evaluation criteria across different observers and offer more clear-cut responses. In this work, we extend the edge of emerging large multi-modality models (LMMs) to further advance visual quality comparison into open-ended settings, that 1) can respond to open-range questions on quality comparison; 2) can provide detailed reasonings beyond direct answers. To this end, we propose the Co-Instruct. To train this first-of-its-kind open-source open-ended visual quality comparer, we collect the Co-Instruct-562K dataset, from two sources: (a) LMM-merged single image quality description, (b) GPT-4V "teacher" responses on unlabeled data. Furthermore, to better evaluate this setting, we propose the MICBench, the first benchmark on multi-image comparison for LMMs. We demonstrate that Co-Instruct not only achieves 30% higher superior accuracy than state-of-the-art open-source LMMs, but also outperforms GPT-4V (its teacher), on both existing related benchmarks and the proposed MICBench. Our model is published at https://huggingface.co/q-future/co-instruct.

link: http://arxiv.org/abs/2402.16641v1

ESG Sentiment Analysis: comparing human and language model performance including GPT

Karim Derrick

In this paper we explore the challenges of measuring sentiment in relation to Environmental, Social and Governance (ESG) social media. ESG has grown in importance in recent years with a surge in interest from the financial sector and the performance of many businesses has become based in part on their ESG related reputations. The use of sentiment analysis to measure ESG related reputation has developed and with it interest in the use of machines to do so. The era of digital media has created an explosion of new media sources, driven by the growth of social media platforms. This growing data environment has become an excellent source for behavioural insight studies across many disciplines that includes politics, healthcare and market research. Our study seeks to compare human performance with the cutting edge in machine performance in the measurement of ESG related sentiment. To this end researchers classify the sentiment of 150 tweets and a reliability measure is made. A gold standard data set is then established based on the consensus of 3 researchers and this data set is then used to measure the performance of different machine approaches: one based on the VADER dictionary approach to sentiment classification and then multiple language model approaches, including Llama2, T5, Mistral, Mixtral, FINBERT, GPT3.5 and GPT4.

link: http://arxiv.org/abs/2402.16650v1

A Comprehensive Survey of Belief Rule Base (BRB) Hybrid Expert system: Bridging Decision Science and Professional Services

Karim Derrick

The Belief Rule Base (BRB) system that adopts a hybrid approach integrating the precision of expert systems with the adaptability of data-driven models. Characterized by its use of if-then rules to accommodate various types of uncertainty through belief degrees, BRB adeptly handles fuzziness, randomness, and ignorance. This semi-quantitative tool excels in processing both numerical data and linguistic knowledge from diverse sources, making it as an indispensable resource in modelling complex nonlinear systems. Notably, BRB's transparent, white-box nature ensures accessibility and clarity for decision-makers and stakeholders, further enhancing its applicability. With its growing adoption in fields ranging from decision-making and reliability evaluation in network security and fault diagnosis, this study aims to explore the evolution and the multifaceted applications of BRB. By analysing its development across different domains, we highlight BRB's potential to revolutionize sectors traditionally resistant to technological disruption, in particular insurance and law.

link: http://arxiv.org/abs/2402.16651v1

GigaPevt: Multimodal Medical Assistant

Pavel Blinov, Konstantin Egorov, Ivan Sviridov, Nikolay Ivanov, Stepan Botman, Evgeniy Tagin, Stepan Kudin, Galina Zubkova, Andrey Savchenko

Building an intelligent and efficient medical assistant is still a challenging AI problem. The major limitation comes from the data modality scarceness, which reduces comprehensive patient perception. This demo paper presents the GigaPevt, the first multimodal medical assistant that combines the dialog capabilities of large language models with specialized medical models. Such an approach shows immediate advantages in dialog quality and metric performance, with a 1.18\% accuracy improvement in the question-answering task.

link: http://arxiv.org/abs/2402.16654v1

Enabling robust sensor network design with data processing and optimization making use of local beehive image and video files

Ephrance Eunice Namugenyi, David Tugume, Augustine Kigwana, Benjamin Rukundo

There is an immediate need for creative ways to improve resource efficiency given the dynamic nature of robust sensor networks and their increasing reliance on data-driven approaches.One key challenge faced is efficiently managing large data files collected from sensor networks for example optimal beehive image and video data files. We offer a revolutionary paradigm that uses cutting-edge edge computing techniques to optimize data transmission and storage in order to meet this problem. Our approach encompasses data compression for images and videos, coupled with a data aggregation technique for numerical data. Specifically, we propose a novel compression algorithm that performs better than the traditional Bzip2, in terms of data compression ratio and throughput. We also designed as an addition a data aggregation algorithm that basically performs very well by reducing on the time to process the overhead of individual data packets there by reducing on the network traffic. A key aspect of our approach is its ability to operate in resource-constrained environments, such as that typically found in a local beehive farm application from where we obtained various datasets. To achieve this, we carefully explore key parameters such as throughput, delay tolerance, compression rate, and data retransmission. This ensures that our approach can meet the unique requirements of robust network management while minimizing the impact on resources. Overall, our study presents and majorly focuses on a holistic solution for optimizing data transmission and processing across robust sensor networks for specifically local beehive image and video data files. Our approach has the potential to significantly improve the efficiency and effectiveness of robust sensor network management, thereby supporting sustainable practices in various IoT applications such as in Bee Hive Data Management.

link: http://arxiv.org/abs/2402.16655v1

Penalized Generative Variable Selection

Tong Wang, Jian Huang, Shuangge Ma

Deep networks are increasingly applied to a wide variety of data, including data with high-dimensional predictors. In such analysis, variable selection can be needed along with estimation/model building. Many of the existing deep network studies that incorporate variable selection have been limited to methodological and numerical developments. In this study, we consider modeling/estimation using the conditional Wasserstein Generative Adversarial networks. Group Lasso penalization is applied for variable selection, which may improve model estimation/prediction, interpretability, stability, etc. Significantly advancing from the existing literature, the analysis of censored survival data is also considered. We establish the convergence rate for variable selection while considering the approximation error, and obtain a more efficient distribution estimation. Simulations and the analysis of real experimental data demonstrate satisfactory practical utility of the proposed analysis.

link: http://arxiv.org/abs/2402.16661v1

UN-SAM: Universal Prompt-Free Segmentation for Generalized Nuclei Images

Zhen Chen, Qing Xu, Xinyu Liu, Yixuan Yuan

In digital pathology, precise nuclei segmentation is pivotal yet challenged by the diversity of tissue types, staining protocols, and imaging conditions. Recently, the segment anything model (SAM) revealed overwhelming performance in natural scenarios and impressive adaptation to medical imaging. Despite these advantages, the reliance of labor-intensive manual annotation as segmentation prompts severely hinders their clinical applicability, especially for nuclei image analysis containing massive cells where dense manual prompts are impractical. To overcome the limitations of current SAM methods while retaining the advantages, we propose the Universal prompt-free SAM framework for Nuclei segmentation (UN-SAM), by providing a fully automated solution with remarkable generalization capabilities. Specifically, to eliminate the labor-intensive requirement of per-nuclei annotations for prompt, we devise a multi-scale Self-Prompt Generation (SPGen) module to revolutionize clinical workflow by automatically generating high-quality mask hints to guide the segmentation tasks. Moreover, to unleash the generalization capability of SAM across a variety of nuclei images, we devise a Domain-adaptive Tuning Encoder (DT-Encoder) to seamlessly harmonize visual features with domain-common and domain-specific knowledge, and further devise a Domain Query-enhanced Decoder (DQ-Decoder) by leveraging learnable domain queries for segmentation decoding in different nuclei domains. Extensive experiments prove that UN-SAM with exceptional performance surpasses state-of-the-arts in nuclei instance and semantic segmentation, especially the generalization capability in zero-shot scenarios. The source code is available at https://github.com/CUHK-AIM-Group/UN-SAM.

link: http://arxiv.org/abs/2402.16663v1

RepoAgent: An LLM-Powered Open-Source Framework for Repository-level Code Documentation Generation

Qinyu Luo, Yining Ye, Shihao Liang, Zhong Zhang, Yujia Qin, Yaxi Lu, Yesai Wu, Xin Cong, Yankai Lin, Yingli Zhang, Xiaoyin Che, Zhiyuan Liu, Maosong Sun

Generative models have demonstrated considerable potential in software engineering, particularly in tasks such as code generation and debugging. However, their utilization in the domain of code documentation generation remains underexplored. To this end, we introduce RepoAgent, a large language model powered open-source framework aimed at proactively generating, maintaining, and updating code documentation. Through both qualitative and quantitative evaluations, we have validated the effectiveness of our approach, showing that RepoAgent excels in generating high-quality repository-level documentation. The code and results are publicly accessible at https://github.com/OpenBMB/RepoAgent.

link: http://arxiv.org/abs/2402.16667v1

Program-Based Strategy Induction for Reinforcement Learning

Carlos G. Correa, Thomas L. Griffiths, Nathaniel D. Daw

Typical models of learning assume incremental estimation of continuously-varying decision variables like expected rewards. However, this class of models fails to capture more idiosyncratic, discrete heuristics and strategies that people and animals appear to exhibit. Despite recent advances in strategy discovery using tools like recurrent networks that generalize the classic models, the resulting strategies are often onerous to interpret, making connections to cognition difficult to establish. We use Bayesian program induction to discover strategies implemented by programs, letting the simplicity of strategies trade off against their effectiveness. Focusing on bandit tasks, we find strategies that are difficult or unexpected with classical incremental learning, like asymmetric learning from rewarded and unrewarded trials, adaptive horizon-dependent random exploration, and discrete state switching.

link: http://arxiv.org/abs/2402.16668v1

Pay Attention: a Call to Regulate the Attention Market and Prevent Algorithmic Emotional Governance

Franck Michel, Fabien Gandon

Over the last 70 years, we, humans, have created an economic market where attention is being captured and turned into money thanks to advertising. During the last two decades, leveraging research in psychology, sociology, neuroscience and other domains, Web platforms have brought the process of capturing attention to an unprecedented scale. With the initial commonplace goal of making targeted advertising more effective, the generalization of attention-capturing techniques and their use of cognitive biases and emotions have multiple detrimental side effects such as polarizing opinions, spreading false information and threatening public health, economies and democracies. This is clearly a case where the Web is not used for the common good and where, in fact, all its users become a vulnerable population. This paper brings together contributions from a wide range of disciplines to analyze current practices and consequences thereof. Through a set of propositions and principles that could be used do drive further works, it calls for actions against these practices competing to capture our attention on the Web, as it would be unsustainable for a civilization to allow attention to be wasted with impunity on a world-wide scale.

link: http://arxiv.org/abs/2402.16670v1