Tue 2024.03.19

MIntRec2.0: A Large-scale Benchmark Dataset for Multimodal Intent Recognition and Out-of-scope Detection in Conversations

Hanlei Zhang, Xin Wang, Hua Xu, Qianrui Zhou, Kai Gao, Jianhua Su, jinyue Zhao, Wenrui Li, Yanting Chen

Multimodal intent recognition poses significant challenges, requiring the incorporation of non-verbal modalities from real-world contexts to enhance the comprehension of human intentions. Existing benchmark datasets are limited in scale and suffer from difficulties in handling out-of-scope samples that arise in multi-turn conversational interactions. We introduce MIntRec2.0, a large-scale benchmark dataset for multimodal intent recognition in multi-party conversations. It contains 1,245 dialogues with 15,040 samples, each annotated within a new intent taxonomy of 30 fine-grained classes. Besides 9,304 in-scope samples, it also includes 5,736 out-of-scope samples appearing in multi-turn contexts, which naturally occur in real-world scenarios. Furthermore, we provide comprehensive information on the speakers in each utterance, enriching its utility for multi-party conversational research. We establish a general framework supporting the organization of single-turn and multi-turn dialogue data, modality feature extraction, multimodal fusion, as well as in-scope classification and out-of-scope detection. Evaluation benchmarks are built using classic multimodal fusion methods, ChatGPT, and human evaluators. While existing methods incorporating nonverbal information yield improvements, effectively leveraging context information and detecting out-of-scope samples remains a substantial challenge. Notably, large language models exhibit a significant performance gap compared to humans, highlighting the limitations of machine learning methods in the cognitive intent understanding task. We believe that MIntRec2.0 will serve as a valuable resource, providing a pioneering foundation for research in human-machine conversational interactions, and significantly facilitating related applications. The full dataset and codes are available at https://github.com/thuiar/MIntRec2.0.

link: http://arxiv.org/abs/2403.10943v1

Human Centered AI for Indian Legal Text Analytics

Sudipto Ghosh, Devanshu Verma, Balaji Ganesan, Purnima Bindal, Vikas Kumar, Vasudha Bhatnagar

Legal research is a crucial task in the practice of law. It requires intense human effort and intellectual prudence to research a legal case and prepare arguments. Recent boom in generative AI has not translated to proportionate rise in impactful legal applications, because of low trustworthiness and and the scarcity of specialized datasets for training Large Language Models (LLMs). This position paper explores the potential of LLMs within Legal Text Analytics (LTA), highlighting specific areas where the integration of human expertise can significantly enhance their performance to match that of experts. We introduce a novel dataset and describe a human centered, compound AI system that principally incorporates human inputs for performing LTA tasks with LLMs.

link: http://arxiv.org/abs/2403.10944v1

The Fallacy of Minimizing Local Regret in the Sequential Task Setting

Ziping Xu, Kelly W. Zhang, Susan A. Murphy

In the realm of Reinforcement Learning (RL), online RL is often conceptualized as an optimization problem, where an algorithm interacts with an unknown environment to minimize cumulative regret. In a stationary setting, strong theoretical guarantees, like a sublinear ($\sqrt{T}$) regret bound, can be obtained, which typically implies the convergence to an optimal policy and the cessation of exploration. However, these theoretical setups often oversimplify the complexities encountered in real-world RL implementations, where tasks arrive sequentially with substantial changes between tasks and the algorithm may not be allowed to adaptively learn within certain tasks. We study the changes beyond the outcome distributions, encompassing changes in the reward designs

(mappings from outcomes to rewards) and the permissible policy spaces. Our results reveal the fallacy of myopically minimizing regret within each task: obtaining optimal regret rates in the early tasks may lead to worse rates in the subsequent ones, even when the outcome distributions stay the same. To realize the optimal cumulative regret bound across all the tasks, the algorithm has to overly explore in the earlier tasks. This theoretical insight is practically significant, suggesting that due to unanticipated changes (e.g., rapid technological development or human-in-the-loop involvement) between tasks, the algorithm needs to explore more than it would in the usual stationary setting within each task. Such implication resonates with the common practice of using clipped policies in mobile health clinical trials and maintaining a fixed rate of ϵ-greedy exploration in robotic learning.

link: http://arxiv.org/abs/2403.10946v1

SelfIE: Self-Interpretation of Large Language Model Embeddings

Haozhe Chen, Carl Vondrick, Chengzhi Mao

How do large language models (LLMs) obtain their answers? The ability to explain and control an LLM's reasoning process is key for reliability, transparency, and future model developments. We propose SelfIE (Self-Interpretation of Embeddings), a framework that enables LLMs to interpret their own embeddings in natural language by leveraging their ability to respond inquiry about a given passage. Capable of interpreting open-world concepts in the hidden embeddings, SelfIE reveals LLM internal reasoning in cases such as making ethical decisions, internalizing prompt injection, and recalling harmful knowledge. SelfIE's text descriptions on hidden embeddings also open up new avenues to control LLM reasoning. We propose Supervised Control, which allows editing open-ended concepts while only requiring gradient computation of individual layer. We extend RLHF to hidden embeddings and propose Reinforcement Control that erases harmful knowledge in LLM without supervision targets.

link: http://arxiv.org/abs/2403.10949v1

Ctrl123: Consistent Novel View Synthesis via Closed-Loop Transcription

Hongxiang Zhao, Xili Dai, Jianan Wang, Shengbang Tong, Jingyuan Zhang, Weida Wang, Lei Zhang, Yi Ma

Large image diffusion models have demonstrated zero-shot capability in novel view synthesis (NVS). However, existing diffusion-based NVS methods struggle to generate novel views that are accurately consistent with the corresponding ground truth poses and appearances, even on the training set. This consequently limits the performance of downstream tasks, such as image-to-multiview generation and 3D reconstruction. We realize that such inconsistency is largely due to the fact that it is difficult to enforce accurate pose and appearance alignment directly in the diffusion training, as mostly done by existing methods such as Zero123. To remedy this problem, we propose Ctrl123, a closed-loop transcription-based NVS diffusion method that enforces alignment between the generated view and ground truth in a pose-sensitive feature space. Our extensive experiments demonstrate the effectiveness of Ctrl123 on the tasks of NVS and 3D reconstruction, achieving significant improvements in both multiview-consistency and pose-consistency over existing methods.

link: http://arxiv.org/abs/2403.10953v1

ClusterSlice: A Zero-touch Deployment Platform for the Edge Cloud Continuum

Lefteris Mamatas, Sotiris Skaperas, Ilias Sakellariou

We demonstrate ClusterSlice, an open-source solution for automated Kubernetes-center deployments for the edge continuum. ClusterSlice is an infrastructure-as-a-service, platform-as-a-service, and application-as-a-service solution, supporting: (i) declarative deployment slice definitions; (ii) infrastructure-on-demand capabilities over multiple heterogeneous domains; (iii) composable Kubernetes deployments, supporting multi-clustering as well as various Kubernetes flavors and intra-cluster/inter-cluster network plugins; (iv) configurable application deployment; and (v) experimentation automation.

link: http://arxiv.org/abs/2403.10954v1

Energy-Based Models with Applications to Speech and Language Processing

Zhijian Ou

Energy-Based Models (EBMs) are an important class of probabilistic models, also known as random fields and undirected graphical models. EBMs are un-normalized and thus radically different from other popular self-normalized probabilistic models such as hidden Markov models (HMMs), autoregressive models, generative adversarial nets (GANs) and variational auto-encoders (VAEs). Over the past years, EBMs have attracted increasing interest not only from the core machine learning community, but also from application domains such as speech, vision, natural language processing (NLP) and so on, due to significant theoretical and algorithmic progress. The sequential nature of speech and language also presents special challenges and needs a different treatment from processing fix-dimensional data (e.g., images). Therefore, the purpose of this monograph is to present a systematic introduction to energy-based models, including both algorithmic progress and applications in speech and language processing. First, the basics of EBMs are introduced, including classic models, recent models parameterized by neural networks, sampling methods, and various learning methods from the classic learning algorithms to the most advanced ones. Then, the application of EBMs in three different scenarios is presented, i.e., for modeling marginal, conditional and joint distributions, respectively. 1) EBMs for sequential data with applications in language modeling, where the main focus is on the marginal distribution of a sequence itself; 2) EBMs for modeling conditional distributions of target sequences given observation sequences, with applications in speech recognition, sequence labeling and text generation; 3) EBMs for modeling joint distributions of both sequences of observations and targets, and their applications in semi-supervised learning and calibrated natural language understanding.

link: http://dx.doi.org/10.1561/2000000117

Exploiting Topological Prior for Boosting Point Cloud Generation

Baiyuan Chen

This paper presents an innovative enhancement to the Sphere as Prior Generative Adversarial Network (SP-GAN) model, a state-of-the-art GAN designed for point cloud generation. A novel method is introduced for point cloud generation that elevates the structural integrity and overall quality of the generated point clouds by incorporating topological priors into the training process of the generator. Specifically, this work utilizes the K-means algorithm to segment a point cloud from the repository into clusters and extract centroids, which are then used as priors in the generation process of the SP-GAN. Furthermore, the discriminator component of the SP-GAN utilizes the identical point cloud that contributed the centroids, ensuring a coherent and consistent learning environment. This strategic use of centroids as intuitive guides not only boosts the efficiency of global feature learning but also substantially improves the structural coherence and fidelity of the generated point clouds. By applying the K-means algorithm to generate centroids as the prior, the work intuitively and experimentally demonstrates that such a prior enhances the quality of generated point clouds.

link: http://arxiv.org/abs/2403.10962v1

Pointer-Generator Networks for Low-Resource Machine Translation: Don't Copy That!

Niyati Bafna, David Yarowsky

While Transformer-based neural machine translation (NMT) is very effective in high-resource settings, many languages lack the necessary large parallel corpora to benefit from it. In the context of low-resource (LR) MT between two closely-related languages, a natural intuition is to seek benefits from structural "shortcuts", such as copying subwords from the source to the target, given that such language pairs often share a considerable number of identical words, cognates, and borrowings. We test Pointer-Generator Networks for this purpose for six language pairs over a variety of resource ranges, and find weak improvements for most settings. However, analysis

shows that the model does not show greater improvements for closely-related vs. more distant language pairs, or for lower resource ranges, and that the models do not exhibit the expected usage of the mechanism for shared subwords. Our discussion of the reasons for this behaviour highlights several general challenges for LR NMT, such as modern tokenization strategies, noisy real-world conditions, and linguistic complexities. We call for better scrutiny of linguistically motivated improvements to NMT given the blackbox nature of Transformer models, as well as for a focus on the above problems in the field.

link: http://arxiv.org/abs/2403.10963v1

Dreaming of Many Worlds: Learning Contextual World Models Aids Zero-Shot Generalization

Sai Prasanna, Karim Farid, Raghu Rajan, André Biedenkapp

Zero-shot generalization (ZSG) to unseen dynamics is a major challenge for creating generally capable embodied agents. To address the broader challenge, we start with the simpler setting of contextual reinforcement learning (cRL), assuming observability of the context values that parameterize the variation in the system's dynamics, such as the mass or dimensions of a robot, without making further simplifying assumptions about the observability of the Markovian state. Toward the goal of ZSG to unseen variation in context, we propose the contextual recurrent state-space model (cRSSM), which introduces changes to the world model of the Dreamer (v3) (Hafner et al., 2023). This allows the world model to incorporate context for inferring latent Markovian states from the observations and modeling the latent dynamics. Our experiments show that such systematic incorporation of the context improves the ZSG of the policies trained on the ``dreams'' of the world model. We further find qualitatively that our approach allows Dreamer to disentangle the latent state from context, allowing it to extrapolate its dreams to the many worlds of unseen contexts. The code for all our experiments is available at \url{https://github.com/sai-prasanna/dreaming_of_many_worlds}.

link: http://arxiv.org/abs/2403.10967v1

Enhancing IoT Security Against DDoS Attacks through Federated Learning

Ghazaleh Shirvani, Saeid Ghasemshirazi, Mohammad Ali Alipour

The rapid proliferation of the Internet of Things (IoT) has ushered in transformative connectivity between physical devices and the digital realm. Nonetheless, the escalating threat of Distributed Denial of Service (DDoS) attacks jeopardizes the integrity and reliability of IoT networks. Conventional DDoS mitigation approaches are ill-equipped to handle the intricacies of IoT ecosystems, potentially compromising data privacy. This paper introduces an innovative strategy to bolster the security of IoT networks against DDoS attacks by harnessing the power of Federated Learning that allows multiple IoT devices or edge nodes to collaboratively build a global model while preserving data privacy and minimizing communication overhead. The research aims to investigate Federated Learning's effectiveness in detecting and mitigating DDoS attacks in IoT. Our proposed framework leverages IoT devices' collective intelligence for real-time attack detection without compromising sensitive data. This study proposes innovative deep autoencoder approaches for data dimensionality reduction, retraining, and partial selection to enhance the performance and stability of the proposed model. Additionally, two renowned aggregation algorithms, FedAvg and FedAvgM, are employed in this research. Various metrics, including true positive rate, false positive rate, and F1-score, are employed to evaluate the model. The dataset utilized in this research, N-BaIoT, exhibits non-IID data distribution, where data categories are distributed quite differently. The negative impact of these distribution disparities is managed by employing retraining and partial selection techniques, enhancing the final model's stability. Furthermore, evaluation results demonstrate that the FedAvgM aggregation algorithm outperforms FedAvg, indicating that in non-IID datasets, FedAvgM provides better stability and performance.

link: http://arxiv.org/abs/2403.10968v1

Task-Aware Low-Rank Adaptation of Segment Anything Model

Xuehao Wang, Feiyang Ye, Yu Zhang

The Segment Anything Model (SAM), with its remarkable zero-shot capability, has been proven to be a powerful foundation model for image segmentation tasks, which is an important task in computer vision. However, the transfer of its rich semantic information to multiple different downstream tasks remains unexplored. In this paper, we propose the Task-Aware Low-Rank Adaptation (TA-LoRA) method, which enables SAM to work as a foundation model for multi-task learning. Specifically, TA-LoRA injects an update parameter tensor into each layer of the encoder in SAM and leverages a low-rank tensor decomposition method to incorporate both task-shared and task-specific information. Furthermore, we introduce modified SAM (mSAM) for multi-task learning where we remove the prompt encoder of SAM and use task-specific no mask embeddings and mask decoder for each task. Extensive experiments conducted on benchmark datasets substantiate the efficacy of TA-LoRA in enhancing the performance of mSAM across multiple downstream tasks.

link: http://arxiv.org/abs/2403.10971v1

An Open-Source Experimentation Framework for the Edge Cloud Continuum

Georgios Koukis, Sotiris Skaperas, Ioanna Angeliki Kapetanidou, Vassilis Tsaoussidis, Lefteris Mamatas

The CODECO Experimentation Framework is an open-source solution designed for the rapid experimentation of Kubernetes-based edge cloud deployments. It adopts a microservice-based architecture and introduces innovative abstractions for (i) the holistic deployment of Kubernetes clusters and associated applications, starting from the VM allocation level; (ii) declarative cross-layer experiment configuration; and (iii) automation features covering the entire experimental process, from the configuration up to the results visualization. We present proof-of-concept results that demonstrate the above capabilities in three distinct contexts: (i) a comparative evaluation of various network fabrics across different edge-oriented Kubernetes distributions; (ii) the automated deployment of EdgeNet, which is a complex edge cloud orchestration system; and (iii) an assessment of anomaly detection (AD) workflows tailored for edge environments.

link: http://arxiv.org/abs/2403.10977v1

Entity Alignment with Unlabeled Dangling Cases

Hang Yin, Dong Ding, Liyao Xiang, Yuheng He, Yihan Wu, Xinbing Wang, Chenghu Zhou

We investigate the entity alignment problem with unlabeled dangling cases, meaning that there are entities in the source or target graph having no counterparts in the other, and those entities remain unlabeled. The problem arises when the source and target graphs are of different scales, and it is much cheaper to label the matchable pairs than the dangling entities. To solve the issue, we propose a novel GNN-based dangling detection and entity alignment framework. While the two tasks share the same GNN and are trained together, the detected dangling entities are removed in the alignment. Our framework is featured by a designed entity and relation attention mechanism for selective neighborhood aggregation in representation learning, as well as a positive-unlabeled learning loss for an unbiased estimation of dangling entities. Experimental results have shown that each component of our design contributes to the overall alignment performance which is comparable or superior to baselines, even if the baselines additionally have 30\% of the dangling entities labeled as training data.

link: http://arxiv.org/abs/2403.10978v1

Automatic Spatial Calibration of Near-Field MIMO Radar With Respect to Optical Sensors

Vanessa Wirth, Johanna Bräunig, Danti Khouri, Florian Gutsche, Martin Vossiek, Tim Weyrich, Marc Stamminger

Despite an emerging interest in MIMO radar, the utilization of its complementary strengths in combination with optical sensors has so far been limited to far-field applications, due to the

challenges that arise from mutual sensor calibration in the near field. In fact, most related approaches in the autonomous industry propose target-based calibration methods using corner reflectors that have proven to be unsuitable for the near field. In contrast, we propose a novel, joint calibration approach for optical RGB-D sensors and MIMO radars that is designed to operate in the radar's near-field range, within decimeters from the sensors. Our pipeline consists of a bespoke calibration target, allowing for automatic target detection and localization, followed by the spatial calibration of the two sensor coordinate systems through target registration. We validate our approach using two different depth sensing technologies from the optical domain. The experiments show the efficiency and accuracy of our calibration for various target displacements, as well as its robustness of our localization in terms of signal ambiguities.

link: http://arxiv.org/abs/2403.10981v1