Fri 2024.04.05

Large Language Models for Orchestrating Bimanual Robots

Kun Chu, Xufeng Zhao, Cornelius Weber, Mengdi Li, Wenhao Lu, Stefan Wermter

Although there has been rapid progress in endowing robots with the ability to solve complex manipulation tasks, generating control policies for bimanual robots to solve tasks involving two hands is still challenging because of the difficulties in effective temporal and spatial coordination. With emergent abilities in terms of step-by-step reasoning and in-context learning, Large Language Models (LLMs) have taken control of a variety of robotic tasks. However, the nature of language communication via a single sequence of discrete symbols makes LLM-based coordination in continuous space a particular challenge for bimanual tasks. To tackle this challenge for the first time by an LLM, we present LAnguage-model-based Bimanual ORchestration (LABOR), an agent utilizing an LLM to analyze task configurations and devise coordination control policies for addressing long-horizon bimanual tasks. In the simulated environment, the LABOR agent is evaluated through several everyday tasks on the NICOL humanoid robot. Reported success rates indicate that overall coordination efficiency is close to optimal performance, while the analysis of failure causes, classified into spatial and temporal coordination and skill selection, shows that these vary over tasks. The project website can be found at http://labor-agent.github.io

link: http://arxiv.org/abs/2404.02018v1

Improving Retrieval Augmented Open-Domain Question-Answering with Vectorized Contexts

Zhuo Chen, Xinyu Wang, Yong Jiang, Pengjun Xie, Fei Huang, Kewei Tu

In the era of large language models, applying techniques such as Retrieval Augmented Generation can better address Open-Domain Question-Answering problems. Due to constraints including model sizes and computing resources, the length of context is often limited, and it becomes challenging to empower the model to cover overlong contexts while answering questions from open domains. This paper proposes a general and convenient method to covering longer contexts in Open-Domain Question-Answering tasks. It leverages a small encoder language model that effectively encodes contexts, and the encoding applies cross-attention with origin inputs. With our method, the origin language models can cover several times longer contexts while keeping the computing requirements close to the baseline. Our experiments demonstrate that after fine-tuning, there is improved performance across two held-in datasets, four held-out datasets, and also in two In Context Learning settings.

link: http://arxiv.org/abs/2404.02022v1

MultiParaDetox: Extending Text Detoxification with Parallel Data to New Languages

Daryna Dementieva, Nikolay Babakov, Alexander Panchenko

Text detoxification is a textual style transfer (TST) task where a text is paraphrased from a toxic surface form, e.g. featuring rude words, to the neutral register. Recently, text detoxification methods found their applications in various task such as detoxification of Large Language Models (LLMs) (Leong et al., 2023; He et al., 2024; Tang et al., 2023) and toxic speech combating in social networks (Deng et al., 2023; Mun et al., 2023; Agarwal et al., 2023). All these applications are extremely important to ensure safe communication in modern digital worlds. However, the previous approaches for parallel text detoxification corpora collection -- ParaDetox (Logacheva et al., 2022) and APPADIA (Atwell et al., 2022) -- were explored only in monolingual setup. In this work, we aim to extend ParaDetox pipeline to multiple languages presenting MultiParaDetox to automate parallel detoxification corpus collection for potentially any language. Then, we experiment with different text detoxification models -- from unsupervised baselines to LLMs and fine-tuned models on the presented parallel corpora -- showing the great benefit of parallel corpus presence to obtain state-of-the-art text detoxification models for any language.

link: http://arxiv.org/abs/2404.02037v1

A Survey on Large Language Model-Based Game Agents

Sihao Hu, Tiansheng Huang, Fatih Ilhan, Selim Tekin, Gaowen Liu, Ramana Kompella, Ling Liu

The development of game agents holds a critical role in advancing towards Artificial General Intelligence (AGI). The progress of LLMs and their multimodal counterparts (MLLMs) offers an unprecedented opportunity to evolve and empower game agents with human-like decision-making capabilities in complex computer game environments. This paper provides a comprehensive overview of LLM-based game agents from a holistic viewpoint. First, we introduce the conceptual architecture of LLM-based game agents, centered around six essential functional components: perception, memory, thinking, role-playing, action, and learning. Second, we survey existing representative LLM-based game agents documented in the literature with respect to methodologies and adaptation agility across six genres of games, including adventure, communication, competition, cooperation, simulation, and crafting & exploration games. Finally, we present an outlook of future research and development directions in this burgeoning field. A curated list of relevant papers is maintained and made accessible at: https://github.com/git-disl/awesome-LLM-game-agent-papers.

link: http://arxiv.org/abs/2404.02039v1

Transformers as Transducers

Lena Strobl, Dana Angluin, David Chiang, Jonathan Rawski, Ashish Sabharwal

We study the sequence-to-sequence mapping capacity of transformers by relating them to finite transducers, and find that they can express surprisingly large classes of transductions. We do so using variants of RASP, a programming language designed to help people "think like transformers," as an intermediate representation. We extend the existing Boolean variant B-RASP to sequence-to-sequence functions and show that it computes exactly the first-order rational functions (such as string rotation). Then, we introduce two new extensions. B-RASP[pos] enables calculations on positions (such as copying the first half of a string) and contains all first-order regular functions. S-RASP adds prefix sum, which enables additional arithmetic operations (such as squaring a string) and contains all first-order polyregular functions. Finally, we show that masked average-hard attention transformers can simulate S-RASP. A corollary of our results is a new proof that transformer decoders are Turing-complete.

link: http://arxiv.org/abs/2404.02040v1

SelfPose3d: Self-Supervised Multi-Person Multi-View 3d Pose Estimation

Vinkle Srivastav, Keqi Chen, Nicolas Padoy

We present a new self-supervised approach, SelfPose3d, for estimating 3d poses of multiple persons from multiple camera views. Unlike current state-of-the-art fully-supervised methods, our approach does not require any 2d or 3d ground-truth poses and uses only the multi-view input images from a calibrated camera setup and 2d pseudo poses generated from an off-the-shelf 2d human pose estimator. We propose two self-supervised learning objectives: self-supervised person localization in 3d space and self-supervised 3d pose estimation. We achieve self-supervised 3d person localization by training the model on synthetically generated 3d points, serving as 3d person root positions, and on the projected root-heatmaps in all the views. We then model the 3d poses of all the localized persons with a bottleneck representation, map them onto all views obtaining 2d joints, and render them using 2d Gaussian heatmaps in an end-to-end differentiable manner. Afterwards, we use the corresponding 2d joints and heatmaps from the pseudo 2d poses for learning. To alleviate the intrinsic inaccuracy of the pseudo labels, we propose an adaptive supervision attention mechanism to guide the self-supervision. Our experiments and analysis on three public benchmark datasets, including Panoptic, Shelf, and Campus, show the effectiveness of our approach, which is comparable to fully-supervised methods. Code is available at \url{https://github.com/CAMMA-public/SelfPose3D}

link: http://arxiv.org/abs/2404.02041v1

Ukrainian Texts Classification: Exploration of Cross-lingual Knowledge Transfer Approaches

Daryna Dementieva, Valeriia Khylenko, Georg Groh

Despite the extensive amount of labeled datasets in the NLP text classification field, the persistent imbalance in data availability across various languages remains evident. Ukrainian, in particular, stands as a language that still can benefit from the continued refinement of cross-lingual methodologies. Due to our knowledge, there is a tremendous lack of Ukrainian corpora for typical text classification tasks. In this work, we leverage the state-of-the-art advances in NLP, exploring cross-lingual knowledge transfer methods avoiding manual data curation: large multilingual encoders and translation systems, LLMs, and language adapters. We test the approaches on three text classification tasks -- toxicity classification, formality classification, and natural language inference -- providing the "recipe" for the optimal setups.

link: http://arxiv.org/abs/2404.02043v1

Causality-based Transfer of Driving Scenarios to Unseen Intersections

Christoph Glasmacher, Michael Schuldes, Sleiman El Masri, Lutz Eckstein

Scenario-based testing of automated driving functions has become a promising method to reduce time and cost compared to real-world testing. In scenario-based testing automated functions are evaluated in a set of pre-defined scenarios. These scenarios provide information about vehicle behaviors, environmental conditions, or road characteristics using parameters. To create realistic scenarios, parameters and parameter dependencies have to be fitted utilizing real-world data. However, due to the large variety of intersections and movement constellations found in reality, data may not be available for certain scenarios. This paper proposes a methodology to systematically analyze relations between parameters of scenarios. Bayesian networks are utilized to analyze causal dependencies in order to decrease the amount of required data and to transfer causal patterns creating unseen scenarios. Thereby, infrastructural influences on movement patterns are investigated to generate realistic scenarios on unobserved intersections. For evaluation, scenarios and underlying parameters are extracted from the inD dataset. Movement patterns are estimated, transferred and checked against recorded data from those initially unseen intersections.

link: http://arxiv.org/abs/2404.02046v1

Universal representations for financial transactional data: embracing local, global, and external contexts

Alexandra Bazarova, Maria Kovaleva, Ilya Kuleshov, Evgenia Romanenkova, Alexander Stepikin, Alexandr Yugay, Dzhambulat Mollaev, Ivan Kireev, Andrey Savchenko, Alexey Zaytsev

Effective processing of financial transactions is essential for banking data analysis. However, in this domain, most methods focus on specialized solutions to stand-alone problems instead of constructing universal representations suitable for many problems. We present a representation learning framework that addresses diverse business challenges. We also suggest novel generative models that account for data specifics, and a way to integrate external information into a client's representation, leveraging insights from other customers' actions. Finally, we offer a benchmark, describing representation quality globally, concerning the entire transaction history; locally, reflecting the client's current state; and dynamically, capturing representation evolution over time. Our generative approach demonstrates superior performance in local tasks, with an increase in ROC-AUC of up to 14\% for the next MCC prediction task and up to 46\% for downstream tasks from existing contrastive baselines. Incorporating external information improves the scores by an additional 20\%.

link: http://arxiv.org/abs/2404.02047v1

Noise Masking Attacks and Defenses for Pretrained Speech Models

Matthew Jagielski, Om Thakkar, Lun Wang

Speech models are often trained on sensitive data in order to improve model performance, leading to potential privacy leakage. Our work considers noise masking attacks, introduced by Amid et al. 2022, which attack automatic speech recognition (ASR) models by requesting a transcript of an utterance which is partially replaced with noise. They show that when a record has been seen at training time, the model will transcribe the noisy record with its memorized sensitive transcript. In our work, we extend these attacks beyond ASR models, to attack pretrained speech encoders. Our method fine-tunes the encoder to produce an ASR model, and then performs noise masking on this model, which we find recovers private information from the pretraining data, despite the model never having seen transcripts at pretraining time! We show how to improve the precision of these attacks and investigate a number of countermeasures to our attacks.

link: http://arxiv.org/abs/2404.02052v1

BERTopic-Driven Stock Market Predictions: Unraveling Sentiment Insights

Enmin Zhu, Jerome Yen

This paper explores the intersection of Natural Language Processing (NLP) and financial analysis, focusing on the impact of sentiment analysis in stock price prediction. We employ BERTopic, an advanced NLP technique, to analyze the sentiment of topics derived from stock market comments. Our methodology integrates this sentiment analysis with various deep learning models, renowned for their effectiveness in time series and stock prediction tasks. Through comprehensive experiments, we demonstrate that incorporating topic sentiment notably enhances the performance of these models. The results indicate that topics in stock market comments provide implicit, valuable insights into stock market volatility and price trends. This study contributes to the field by showcasing the potential of NLP in enriching financial analysis and opens up avenues for further research into real-time sentiment analysis and the exploration of emotional and contextual aspects of market sentiment. The integration of advanced NLP techniques like BERTopic with traditional financial analysis methods marks a step forward in developing more sophisticated tools for understanding and predicting market behaviors.

link: http://arxiv.org/abs/2404.02053v2

Deconstructing In-Context Learning: Understanding Prompts via Corruption

Namrata Shivagunde, Vladislav Lialin, Sherin Muckatira, Anna Rumshisky

The ability of large language models (LLMs) to "learn in context" based on the provided prompt has led to an explosive growth in their use, culminating in the proliferation of AI assistants such as ChatGPT, Claude, and Bard. These AI assistants are known to be robust to minor prompt modifications, mostly due to alignment techniques that use human feedback. In contrast, the underlying pre-trained LLMs they use as a backbone are known to be brittle in this respect. Building high-quality backbone models remains a core challenge, and a common approach to assessing their quality is to conduct few-shot evaluation. Such evaluation is notorious for being highly sensitive to minor prompt modifications, as well as the choice of specific in-context examples. Prior work has examined how modifying different elements of the prompt can affect model performance. However, these earlier studies tended to concentrate on a limited number of specific prompt attributes and often produced contradictory results. Additionally, previous research either focused on models with fewer than 15 billion parameters or exclusively examined black-box models like GPT-3 or PaLM, making replication challenging. In the present study, we decompose the entire prompt into four components: task description, demonstration inputs, labels, and inline instructions provided for each demonstration. We investigate the effects of structural and semantic corruptions of these elements on model performance. We study models ranging from 1.5B to 70B in size, using ten datasets covering classification and generation tasks. We find that repeating text within the prompt boosts model performance, and bigger models ($\geq$30B) are more sensitive to the semantics of the prompt. Finally, we observe that adding task and inline instructions to the demonstrations enhances model performance even when the instructions are semantically corrupted.

link: http://arxiv.org/abs/2404.02054v1

IISAN: Efficiently Adapting Multimodal Representation for Sequential Recommendation with Decoupled PEFT

Junchen Fu, Xuri Ge, Xin Xin, Alexandros Karatzoglou, Ioannis Arapakis, Jie Wang, Joemon M Jose

Multimodal foundation models are transformative in sequential recommender systems, leveraging powerful representation learning capabilities. While Parameter-efficient Fine-tuning (PEFT) is commonly used to adapt foundation models for recommendation tasks, most research prioritizes parameter efficiency, often overlooking critical factors like GPU memory efficiency and training speed. Addressing this gap, our paper introduces IISAN (Intra- and Inter-modal Side Adapted Network for Multimodal Representation), a simple plug-and-play architecture using a Decoupled PEFT structure and exploiting both intra- and inter-modal adaptation. IISAN matches the performance of full fine-tuning (FFT) and state-of-the-art PEFT. More importantly, it significantly reduces GPU memory usage - from 47GB to just 3GB for multimodal sequential recommendation tasks. Additionally, it accelerates training time per epoch from 443s to 22s compared to FFT. This is also a notable improvement over the Adapter and LoRA, which require 37-39 GB GPU memory and 350-380 seconds per epoch for training. Furthermore, we propose a new composite efficiency metric, TPME (Training-time, Parameter, and GPU Memory Efficiency) to alleviate the prevalent misconception that "parameter efficiency represents overall efficiency". TPME provides more comprehensive insights into practical efficiency comparisons between different methods. Besides, we give an accessible efficiency analysis of all PEFT and FFT approaches, which demonstrate the superiority of IISAN. We release our codes and other materials at https://github.com/jjGenAILab/IISAN.

link: http://arxiv.org/abs/2404.02059v1

Long-context LLMs Struggle with Long In-context Learning

Tianle Li, Ge Zhang, Quy Duc Do, Xiang Yue, Wenhu Chen

Large Language Models (LLMs) have made significant strides in handling long sequences exceeding 32K tokens. However, their performance evaluation has largely been confined to metrics like perplexity and synthetic tasks, which may not fully capture their abilities in more nuanced, real-world scenarios. This study introduces a specialized benchmark (LongICLBench) focusing on long in-context learning within the realm of extreme-label classification. We meticulously selected six datasets with a label range spanning 28 to 174 classes covering different input (few-shot demonstration) lengths from 2K to 50K tokens. Our benchmark requires LLMs to comprehend the entire input to recognize the massive label spaces to make correct predictions. We evaluate 13 long-context LLMs on our benchmarks. We find that the long-context LLMs perform relatively well on less challenging tasks with shorter demonstration lengths by effectively utilizing the long context window. However, on the most challenging task Discovery with 174 labels, all the LLMs struggle to understand the task definition, thus reaching a performance close to zero. This suggests a notable gap in current LLM capabilities for processing and understanding long, context-rich sequences. Further analysis revealed a tendency among models to favor predictions for labels presented toward the end of the sequence. Their ability to reason over multiple pieces in the long sequence is yet to be improved. Our study reveals that long context understanding and reasoning is still a challenging task for the existing LLMs. We believe LongICLBench could serve as a more realistic evaluation for the future long-context LLMs.

link: http://arxiv.org/abs/2404.02060v2

Digital Forgetting in Large Language Models: A Survey of Unlearning Methods

Alberto Blanco-Justicia, Najeeb Jebreel, Benet Manzanares, David Sánchez, Josep Domingo-Ferrer, Guillem Collell, Kuan Eeik Tan

The objective of digital forgetting is, given a model with undesirable knowledge or behavior, obtain a new model where the detected issues are no longer present. The motivations for forgetting include

privacy protection, copyright protection, elimination of biases and discrimination, and prevention of harmful content generation. Effective digital forgetting has to be effective (meaning how well the new model has forgotten the undesired knowledge/behavior), retain the performance of the original model on the desirable tasks, and be scalable (in particular forgetting has to be more efficient than retraining from scratch on just the tasks/data to be retained). This survey focuses on forgetting in large language models (LLMs). We first provide background on LLMs, including their components, the types of LLMs, and their usual training pipeline. Second, we describe the motivations, types, and desired properties of digital forgetting. Third, we introduce the approaches to digital forgetting in LLMs, among which unlearning methodologies stand out as the state of the art. Fourth, we provide a detailed taxonomy of machine unlearning methods for LLMs, and we survey and compare current approaches. Fifth, we detail datasets, models and metrics used for the evaluation of forgetting, retaining and runtime. Sixth, we discuss challenges in the area. Finally, we provide some concluding remarks.

link: http://arxiv.org/abs/2404.02062v1