

Fri 2024.04.12

SurveyAgent: A Conversational System for Personalized and Efficient Research Survey

Xintao Wang, Jiangjie Chen, Nianqi Li, Lida Chen, Xinfeng Yuan, Wei Shi, Xuyang Ge, Rui Xu, Yanghua Xiao

In the rapidly advancing research fields such as AI, managing and staying abreast of the latest scientific literature has become a significant challenge for researchers. Although previous efforts have leveraged AI to assist with literature searches, paper recommendations, and question-answering, a comprehensive support system that addresses the holistic needs of researchers has been lacking. This paper introduces SurveyAgent, a novel conversational system designed to provide personalized and efficient research survey assistance to researchers. SurveyAgent integrates three key modules: Knowledge Management for organizing papers, Recommendation for discovering relevant literature, and Query Answering for engaging with content on a deeper level. This system stands out by offering a unified platform that supports researchers through various stages of their literature review process, facilitated by a conversational interface that prioritizes user interaction and personalization. Our evaluation demonstrates SurveyAgent's effectiveness in streamlining research activities, showcasing its capability to facilitate how researchers interact with scientific literature.

link: <http://arxiv.org/abs/2404.06364v1>

Dynamic Resolution Guidance for Facial Expression Recognition

Jie Ou, Xu Li, Tianxiang Jiang, Yuanlun Xie

Facial expression recognition (FER) is vital for human-computer interaction and emotion analysis, yet recognizing expressions in low-resolution images remains challenging. This paper introduces a practical method called Dynamic Resolution Guidance for Facial Expression Recognition (DRGFER) to effectively recognize facial expressions in images with varying resolutions without compromising FER model accuracy. Our framework comprises two main components: the Resolution Recognition Network (RRN) and the Multi-Resolution Adaptation Facial Expression Recognition Network (MRAFER). The RRN determines image resolution, outputs a binary vector, and the MRAFER assigns images to suitable facial expression recognition networks based on resolution. We evaluated DRGFER on widely-used datasets RAFDB and FERPlus, demonstrating that our method retains optimal model performance at each resolution and outperforms alternative resolution approaches. The proposed framework exhibits robustness against resolution variations and facial expressions, offering a promising solution for real-world applications.

link: <http://arxiv.org/abs/2404.06365v1>

ClinLinker: Medical Entity Linking of Clinical Concept Mentions in Spanish

Fernando Gallego, Guillermo López-García, Luis Gasco-Sánchez, Martin Krallinger, Francisco J. Veredas

Advances in natural language processing techniques, such as named entity recognition and normalization to widely used standardized terminologies like UMLS or SNOMED-CT, along with the digitalization of electronic health records, have significantly advanced clinical text analysis. This study presents ClinLinker, a novel approach employing a two-phase pipeline for medical entity linking that leverages the potential of in-domain adapted language models for biomedical text mining: initial candidate retrieval using a SapBERT-based bi-encoder and subsequent re-ranking with a cross-encoder, trained by following a contrastive-learning strategy to be tailored to medical concepts in Spanish. This methodology, focused initially on content in Spanish, substantially outperforming multilingual language models designed for the same purpose. This is true even for complex scenarios involving heterogeneous medical terminologies and being trained on a subset of the original data. Our results, evaluated using top-k accuracy at 25 and other top-k metrics, demonstrate our approach's performance on two distinct clinical entity linking Gold Standard

corpora, DisTEMIST (diseases) and MedProcNER (clinical procedures), outperforming previous benchmarks by 40 points in DisTEMIST and 43 points in MedProcNER, both normalized to SNOMED-CT codes. These findings highlight our approach's ability to address language-specific nuances and set a new benchmark in entity linking, offering a potent tool for enhancing the utility of digital medical records. The resulting system is of practical value, both for large scale automatic generation of structured data derived from clinical records, as well as for exhaustive extraction and harmonization of predefined clinical variables of interest.

link: <http://arxiv.org/abs/2404.06367v1>

VISION2UI: A Real-World Dataset with Layout for Code Generation from UI Designs

Yi Gui, Zhen Li, Yao Wan, Yemin Shi, Hongyu Zhang, Yi Su, Shaoling Dong, Xing Zhou, Wenbin Jiang

Automatically generating UI code from webpage design visions can significantly alleviate the burden of developers, enabling beginner developers or designers to directly generate Web pages from design diagrams. Currently, prior research has accomplished the objective of generating UI code from rudimentary design visions or sketches through designing deep neural networks. Inspired by the groundbreaking advancements achieved by Multimodal Large Language Models (MLLMs), the automatic generation of UI code from high-fidelity design images is now emerging as a viable possibility. Nevertheless, our investigation reveals that existing MLLMs are hampered by the scarcity of authentic, high-quality, and large-scale datasets, leading to unsatisfactory performance in automated UI code generation. To mitigate this gap, we present a novel dataset, termed VISION2UI, extracted from real-world scenarios, augmented with comprehensive layout information, tailored specifically for finetuning MLLMs in UI code generation. Specifically, this dataset is derived through a series of operations, encompassing collecting, cleaning, and filtering of the open-source Common Crawl dataset. In order to uphold its quality, a neural scorer trained on labeled samples is utilized to refine the data, retaining higher-quality instances. Ultimately, this process yields a dataset comprising 2,000 (Much more is coming soon) parallel samples encompassing design visions and UI code. The dataset is available at <https://huggingface.co/datasets/xcodemind/vision2ui>.

link: <http://arxiv.org/abs/2404.06369v1>

Enhancing Decision Analysis with a Large Language Model: pyDecision a Comprehensive Library of MCDA Methods in Python

Valdecy Pereira, Marcio Pereira Babilio, Carlos Henrique Tarjano Santos Carlos Henrique Tarjano Santos

Purpose: Multicriteria decision analysis (MCDA) has become increasingly essential for decision-making in complex environments. In response to this need, the pyDecision library, implemented in Python and available at <https://bit.ly/3tLFGtH>, has been developed to provide a comprehensive and accessible collection of MCDA methods. **Methods:** The pyDecision offers 70 MCDA methods, including AHP, TOPSIS, and the PROMETHEE and ELECTRE families. Beyond offering a vast range of techniques, the library provides visualization tools for more intuitive results interpretation. In addition to these features, pyDecision has integrated ChatGPT, an advanced Large Language Model, where decision-makers can use ChatGPT to discuss and compare the outcomes of different methods, providing a more interactive and intuitive understanding of the solutions. **Findings:** Large Language Models are undeniably potent but can sometimes be a double-edged sword. Its answers may be misleading without rigorous verification of its outputs, especially for researchers lacking deep domain expertise. It's imperative to approach its insights with a discerning eye and a solid foundation in the relevant field. **Originality:** With the integration of MCDA methods and ChatGPT, pyDecision is a significant contribution to the scientific community, as it is an invaluable resource for researchers, practitioners, and decision-makers navigating complex decision-making problems and seeking the most appropriate solutions based on MCDA methods.

link: <http://arxiv.org/abs/2404.06370v1>

Model Generation from Requirements with LLMs: an Exploratory Study

Alessio Ferrari, Sallam Abualhaija, Chetan Arora

Complementing natural language (NL) requirements with graphical models can improve stakeholders' communication and provide directions for system design. However, creating models from requirements involves manual effort. The advent of generative large language models (LLMs), ChatGPT being a notable example, offers promising avenues for automated assistance in model generation. This paper investigates the capability of ChatGPT to generate a specific type of model, i.e., UML sequence diagrams, from NL requirements. We conduct a qualitative study in which we examine the sequence diagrams generated by ChatGPT for 28 requirements documents of various types and from different domains. Observations from the analysis of the generated diagrams have systematically been captured through evaluation logs, and categorized through thematic analysis. Our results indicate that, although the models generally conform to the standard and exhibit a reasonable level of understandability, their completeness and correctness with respect to the specified requirements often present challenges. This issue is particularly pronounced in the presence of requirements smells, such as ambiguity and inconsistency. The insights derived from this study can influence the practical utilization of LLMs in the RE process, and open the door to novel RE-specific prompting strategies targeting effective model generation.

link: <http://arxiv.org/abs/2404.06371v1>

The Power in Communication: Power Regularization of Communication for Autonomy in Cooperative Multi-Agent Reinforcement Learning

Nancirose Piazza, Vahid Behzadan, Stefan Sarkadi

Communication plays a vital role for coordination in Multi-Agent Reinforcement Learning (MARL) systems. However, misaligned agents can exploit other agents' trust and delegated power to the communication medium. In this paper, we propose power regularization as a method to limit the adverse effects of communication by misaligned agents, specifically communication which impairs the performance of cooperative agents. Power is a measure of the influence one agent's actions have over another agent's policy. By introducing power regularization, we aim to allow designers to control or reduce agents' dependency on communication when appropriate, and make them more resilient to performance deterioration due to misuses of communication. We investigate several environments in which power regularization can be a valuable capability for learning different policies that reduce the effect of power dynamics between agents during communication.

link: <http://arxiv.org/abs/2404.06387v1>

Raster Forge: Interactive Raster Manipulation Library and GUI for Python

Afonso Oliveira, Nuno Fachada, João P. Matos-Carvalho

Raster Forge is a Python library and graphical user interface for raster data manipulation and analysis. The tool is focused on remote sensing applications, particularly in wildfire management. It allows users to import, visualize, and process raster layers for tasks such as image compositing or topographical analysis. For wildfire management, it generates fuel maps using predefined models. Its impact extends from disaster management to hydrological modeling, agriculture, and environmental monitoring. Raster Forge can be a valuable asset for geoscientists and researchers who rely on raster data analysis, enhancing geospatial data processing and visualization across various disciplines.

link: <http://arxiv.org/abs/2404.06389v1>

Latent Distance Guided Alignment Training for Large Language Models

Haotian Luo, Wenhao Zheng, Huaxiu Yao

Ensuring alignment with human preferences is a crucial characteristic of large language models (LLMs). Presently, the primary alignment methods, RLHF and DPO, require extensive human annotation, which is expensive despite their efficacy. The significant expenses associated with

current alignment techniques motivate researchers to investigate the development of annotation-free alignment training methods. In pursuit of improved alignment without relying on external annotation, we introduce Latent Distance Guided Alignment Training (LD-Align). This approach seeks to align the model with a high-quality supervised fine-tune dataset using guidance from a latent space. The latent space is generated through sample reconstruction, akin to auto-encoding. Consequently, we utilize the distance between sample pairs in the latent space to guide DPO-based alignment training. Extensive experimentation and evaluation show the efficacy of our proposed method in achieving notable alignment.

link: <http://arxiv.org/abs/2404.06390v1>

Exploring Neural Network Landscapes: Star-Shaped and Geodesic Connectivity

Zhanran Lin, Puheng Li, Lei Wu

One of the most intriguing findings in the structure of neural network landscape is the phenomenon of mode connectivity: For two typical global minima, there exists a path connecting them without barrier. This concept of mode connectivity has played a crucial role in understanding important phenomena in deep learning. In this paper, we conduct a fine-grained analysis of this connectivity phenomenon. First, we demonstrate that in the overparameterized case, the connecting path can be as simple as a two-piece linear path, and the path length can be nearly equal to the Euclidean distance. This finding suggests that the landscape should be nearly convex in a certain sense. Second, we uncover a surprising star-shaped connectivity: For a finite number of typical minima, there exists a center on minima manifold that connects all of them simultaneously via linear paths. These results are provably valid for linear networks and two-layer ReLU networks under a teacher-student setup, and are empirically supported by models trained on MNIST and CIFAR-10.

link: <http://arxiv.org/abs/2404.06391v1>

Event Extraction in Basque: Typologically motivated Cross-Lingual Transfer-Learning Analysis

Mikel Zubillaga, Oscar Sainz, Ainara Estarrona, Oier Lopez de Lacalle, Eneko Agirre

Cross-lingual transfer-learning is widely used in Event Extraction for low-resource languages and involves a Multilingual Language Model that is trained in a source language and applied to the target language. This paper studies whether the typological similarity between source and target languages impacts the performance of cross-lingual transfer, an under-explored topic. We first focus on Basque as the target language, which is an ideal target language because it is typologically different from surrounding languages. Our experiments on three Event Extraction tasks show that the shared linguistic characteristic between source and target languages does have an impact on transfer quality. Further analysis of 72 language pairs reveals that for tasks that involve token classification such as entity and event trigger identification, common writing script and morphological features produce higher quality cross-lingual transfer. In contrast, for tasks involving structural prediction like argument extraction, common word order is the most relevant feature. In addition, we show that when increasing the training size, not all the languages scale in the same way in the cross-lingual setting. To perform the experiments we introduce EusIE, an event extraction dataset for Basque, which follows the Multilingual Event Extraction dataset (MEE). The dataset and code are publicly available.

link: <http://arxiv.org/abs/2404.06392v1>

MuPT: A Generative Symbolic Music Pretrained Transformer

Xingwei Qu, Yuelin Bai, Yinghao Ma, Ziya Zhou, Ka Man Lo, Jiaheng Liu, Ruibin Yuan, Lejun Min, Xueling Liu, Tianyu Zhang, Xinrun Du, Shuyue Guo, Yiming Liang, Yizhi Li, Shangda Wu, Juntong Zhou, Tianyu Zheng, Ziyang Ma, Fengze Han, Wei Xue, Gus Xia, Emmanouil Benetos, Xiang Yue, Chenghua Lin, Xu Tan, Stephen W. Huang, Wenhui Chen, Jie Fu, Ge Zhang

In this paper, we explore the application of Large Language Models (LLMs) to the pre-training of music. While the prevalent use of MIDI in music modeling is well-established, our findings suggest that LLMs are inherently more compatible with ABC Notation, which aligns more closely with their

design and strengths, thereby enhancing the model's performance in musical composition. To address the challenges associated with misaligned measures from different tracks during generation, we propose the development of a Synchronized Multi-Track ABC Notation (SMT-ABC Notation), which aims to preserve coherence across multiple musical tracks. Our contributions include a series of models capable of handling up to 8192 tokens, covering 90% of the symbolic music data in our training set. Furthermore, we explore the implications of the Symbolic Music Scaling Law (SMS Law) on model performance. The results indicate a promising direction for future research in music generation, offering extensive resources for community-led research through our open-source contributions.

link: <http://arxiv.org/abs/2404.06393v2>

MiniCPM: Unveiling the Potential of Small Language Models with Scalable Training Strategies

Shengding Hu, Yuge Tu, Xu Han, Chaoqun He, Ganqu Cui, Xiang Long, Zhi Zheng, Yewei Fang, Yuxiang Huang, Weilin Zhao, Xinrong Zhang, Zheng Leng Thai, Kaihuo Zhang, Chongyi Wang, Yuan Yao, Chenyang Zhao, Jie Zhou, Jie Cai, Zhongwu Zhai, Ning Ding, Chao Jia, Guoyang Zeng, Dahai Li, Zhiyuan Liu, Maosong Sun

The burgeoning interest in developing Large Language Models (LLMs) with up to trillion parameters has been met with concerns regarding resource efficiency and practical expense, particularly given the immense cost of experimentation. This scenario underscores the importance of exploring the potential of Small Language Models (SLMs) as a resource-efficient alternative. In this context, we introduce MiniCPM, specifically the 1.2B and 2.4B non-embedding parameter variants, not only excel in their respective categories but also demonstrate capabilities on par with 7B-13B LLMs. While focusing on SLMs, our approach exhibits scalability in both model and data dimensions for future LLM research. Regarding model scaling, we employ extensive model wind tunnel experiments for stable and optimal scaling. For data scaling, we introduce a Warmup-Stable-Decay (WSD) learning rate scheduler (LRS), conducive to continuous training and domain adaptation. We present an in-depth analysis of the intriguing training dynamics that occurred in the WSD LRS. With WSD LRS, we are now able to efficiently study data-model scaling law without extensive retraining experiments on both axes of model and data, from which we derive the much higher compute optimal data-model ratio than Chinchilla Optimal. Additionally, we introduce MiniCPM family, including MiniCPM-DPO, MiniCPM-MoE and MiniCPM-128K, whose excellent performance further cementing MiniCPM's foundation in diverse SLM applications. MiniCPM models are available publicly at <https://github.com/OpenBMB/MiniCPM>.

link: <http://arxiv.org/abs/2404.06395v1>

Online Learning of Decision Trees with Thompson Sampling

Ayman Chaouki, Jesse Read, Albert Bifet

Decision Trees are prominent prediction models for interpretable Machine Learning. They have been thoroughly researched, mostly in the batch setting with a fixed labelled dataset, leading to popular algorithms such as C4.5, ID3 and CART. Unfortunately, these methods are of heuristic nature, they rely on greedy splits offering no guarantees of global optimality and often leading to unnecessarily complex and hard-to-interpret Decision Trees. Recent breakthroughs addressed this suboptimality issue in the batch setting, but no such work has considered the online setting with data arriving in a stream. To this end, we devise a new Monte Carlo Tree Search algorithm, Thompson Sampling Decision Trees (TSDT), able to produce optimal Decision Trees in an online setting. We analyse our algorithm and prove its almost sure convergence to the optimal tree. Furthermore, we conduct extensive experiments to validate our findings empirically. The proposed TSDT outperforms existing algorithms on several benchmarks, all while presenting the practical advantage of being tailored to the online setting.

link: <http://arxiv.org/abs/2404.06403v1>

Apprentices to Research Assistants: Advancing Research with Large Language Models

M. Namvarpour, A. Razi

Large Language Models (LLMs) have emerged as powerful tools in various research domains. This article examines their potential through a literature review and firsthand experimentation. While LLMs offer benefits like cost-effectiveness and efficiency, challenges such as prompt tuning, biases, and subjectivity must be addressed. The study presents insights from experiments utilizing LLMs for qualitative analysis, highlighting successes and limitations. Additionally, it discusses strategies for mitigating challenges, such as prompt optimization techniques and leveraging human expertise. This study aligns with the 'LLMs as Research Tools' workshop's focus on integrating LLMs into HCI data work critically and ethically. By addressing both opportunities and challenges, our work contributes to the ongoing dialogue on their responsible application in research.

link: <http://arxiv.org/abs/2404.06404v1>