Sun 2024.03.24

Prediction of Translation Techniques for the Translation Process

Fan Zhou, Vincent Vandeghinste

Machine translation (MT) encompasses a variety of methodologies aimed at enhancing the accuracy of translations. In contrast, the process of human-generated translation relies on a wide range of translation techniques, which are crucial for ensuring linguistic adequacy and fluency. This study suggests that these translation techniques could further optimize machine translation if they are automatically identified before being applied to guide the translation process effectively. The study differentiates between two scenarios of the translation process: from-scratch translation and post-editing. For each scenario, a specific set of experiments has been designed to forecast the most appropriate translation techniques. The findings indicate that the predictive accuracy for from-scratch translation reaches 82%, while the post-editing process exhibits even greater potential, achieving an accuracy rate of 93%.

link: http://arxiv.org/abs/2403.14454v1

gTBLS: Generating Tables from Text by Conditional Question Answering

Anirudh Sundar, Christopher Richardson, Larry Heck

Distilling large, unstructured text into a structured, condensed form such as tables is an open research problem. One of the primary challenges in automatically generating tables is ensuring their syntactic validity. Prior approaches address this challenge by including additional parameters in the Transformer's attention mechanism to attend to specific rows and column headers. In contrast to this single-stage method, this paper presents a two-stage approach called Generative Tables (gTBLS). The first stage infers table structure (row and column headers) from the text. The second stage formulates questions using these headers and fine-tunes a causal language model to answer them. Furthermore, the gTBLS approach is amenable to the utilization of pre-trained Large Language Models in a zero-shot configuration, presenting a solution for table generation in situations where fine-tuning is not feasible. gTBLS improves prior approaches by up to 10% in BERTScore on the table construction task and up to 20% on the table content generation task of the E2E, WikiTableText, WikiBio, and RotoWire datasets.

link: http://arxiv.org/abs/2403.14457v1

Multi-Level Explanations for Generative Language Models

Lucas Monteiro Paes, Dennis Wei, Hyo Jin Do, Hendrik Strobelt, Ronny Luss, Amit Dhurandhar, Manish Nagireddy, Karthikeyan Natesan Ramamurthy, Prasanna Sattigeri, Werner Geyer, Soumya Ghosh

Perturbation-based explanation methods such as LIME and SHAP are commonly applied to text classification. This work focuses on their extension to generative language models. To address the challenges of text as output and long text inputs, we propose a general framework called MExGen that can be instantiated with different attribution algorithms. To handle text output, we introduce the notion of scalarizers for mapping text to real numbers and investigate multiple possibilities. To handle long inputs, we take a multi-level approach, proceeding from coarser levels of granularity to finer ones, and focus on algorithms with linear scaling in model queries. We conduct a systematic evaluation, both automated and human, of perturbation-based attribution methods for summarization and context-grounded question answering. The results show that our framework can provide more locally faithful explanations of generated outputs.

link: http://arxiv.org/abs/2403.14459v1

Towards Single-System Illusion in Software-Defined Vehicles -- Automated, AI-Powered Workflow

Krzysztof Lebioda, Viktor Vorobev, Nenad Petrovic, Fengjunjie Pan, Vahid Zolfaghari, Alois Knoll

We propose a novel model- and feature-based approach to development of vehicle software systems, where the end architecture is not explicitly defined. Instead, it emerges from an iterative process of search and optimization given certain constraints, requirements and hardware architecture, while retaining the property of single-system illusion, where applications run in a logically uniform environment. One of the key points of the presented approach is the inclusion of modern generative AI, specifically Large Language Models (LLMs), in the loop. With the recent advances in the field, we expect that the LLMs will be able to assist in processing of requirements, generation of formal system models, as well as generation of software deployment specification and test code. The resulting pipeline is automated to a large extent, with feedback being generated at each step.

link: http://arxiv.org/abs/2403.14460v1

CathFlow: Self-Supervised Segmentation of Catheters in Interventional Ultrasound Using Optical Flow and Transformers

Alex Ranne, Liming Kuang, Yordanka Velikova, Nassir Navab, Ferdinando Rodriguez y Baena

In minimally invasive endovascular procedures, contrast-enhanced angiography remains the most robust imaging technique. However, it is at the expense of the patient and clinician's health due to prolonged radiation exposure. As an alternative, interventional ultrasound has notable benefits such as being radiation-free, fast to deploy, and having a small footprint in the operating room. Yet, ultrasound is hard to interpret, and highly prone to artifacts and noise. Additionally, interventional radiologists must undergo extensive training before they become qualified to diagnose and treat patients effectively, leading to a shortage of staff, and a lack of open-source datasets. In this work, we seek to address both problems by introducing a self-supervised deep learning architecture to segment catheters in longitudinal ultrasound images, without demanding any labeled data. The network architecture builds upon AiAReSeg, a segmentation transformer built with the Attention in Attention mechanism, and is capable of learning feature changes across time and space. To facilitate training, we used synthetic ultrasound data based on physics-driven catheter insertion simulations, and translated the data into a unique CT-Ultrasound common domain, CACTUSS, to improve the segmentation performance. We generated ground truth segmentation masks by computing the optical flow between adjacent frames using FlowNet2, and performed thresholding to obtain a binary map estimate. Finally, we validated our model on a test dataset, consisting of unseen synthetic data and images collected from silicon aorta phantoms, thus demonstrating its potential for applications to clinical data in the future.

link: http://arxiv.org/abs/2403.14465v1

Universal Feature Selection for Simultaneous Interpretability of Multitask Datasets

Matt Raymond, Jacob Charles Saldinger, Paolo Elvati, Clayton Scott, Angela Violi

Extracting meaningful features from complex, high-dimensional datasets across scientific domains remains challenging. Current methods often struggle with scalability, limiting their applicability to large datasets, or make restrictive assumptions about feature-property relationships, hindering their ability to capture complex interactions. BoUTS's general and scalable feature selection algorithm surpasses these limitations to identify both universal features relevant to all datasets and task-specific features predictive for specific subsets. Evaluated on seven diverse chemical regression datasets, BoUTS achieves state-of-the-art feature sparsity while maintaining prediction accuracy comparable to specialized methods. Notably, BoUTS's universal features enable domain-specific knowledge transfer between datasets, and suggest deep connections in seemingly-disparate chemical datasets. We expect these results to have important repercussions in manually-guided inverse problems. Beyond its current application, BoUTS holds immense potential for elucidating data-poor systems by leveraging information from similar data-rich systems. BoUTS represents a significant leap in cross-domain feature selection, potentially leading to advancements in various scientific fields.

link: http://arxiv.org/abs/2403.14466v1

Recourse for reclamation: Chatting with generative language models

Jennifer Chien, Kevin R. McKee, Jackie Kay, William Isaac

Researchers and developers increasingly rely on toxicity scoring to moderate generative language model outputs, in settings such as customer service, information retrieval, and content generation. However, toxicity scoring may render pertinent information inaccessible, rigidify or "value-lock" cultural norms, and prevent language reclamation processes, particularly for marginalized people. In this work, we extend the concept of algorithmic recourse to generative language models: we provide users a novel mechanism to achieve their desired prediction by dynamically setting thresholds for toxicity filtering. Users thereby exercise increased agency relative to interactions with the baseline system. A pilot study ($n = 30$) supports the potential of our proposed recourse mechanism, indicating improvements in usability compared to fixed-threshold toxicity-filtering of model outputs. Future work should explore the intersection of toxicity scoring, model controllability, user agency, and language reclamation processes -- particularly with regard to the bias that many communities encounter when interacting with generative language models.

link: http://dx.doi.org/10.1145/3613905.3650999

AnyV2V: A Plug-and-Play Framework For Any Video-to-Video Editing Tasks

Max Ku, Cong Wei, Weiming Ren, Huan Yang, Wenhu Chen

Video-to-video editing involves editing a source video along with additional control (such as text prompts, subjects, or styles) to generate a new video that aligns with the source video and the provided control. Traditional methods have been constrained to certain editing types, limiting their ability to meet the wide range of user demands. In this paper, we introduce AnyV2V, a novel training-free framework designed to simplify video editing into two primary steps: (1) employing an off-the-shelf image editing model (e.g. InstructPix2Pix, InstantID, etc) to modify the first frame, (2) utilizing an existing image-to-video generation model (e.g. I2VGen-XL) for DDIM inversion and feature injection. In the first stage, AnyV2V can plug in any existing image editing tools to support an extensive array of video editing tasks. Beyond the traditional prompt-based editing methods, AnyV2V also can support novel video editing tasks, including reference-based style transfer, subject-driven editing, and identity manipulation, which were unattainable by previous methods. In the second stage, AnyV2V can plug in any existing image-to-video models to perform DDIM inversion and intermediate feature injection to maintain the appearance and motion consistency with the source video. On the prompt-based editing, we show that AnyV2V can outperform the previous best approach by 35\% on prompt alignment, and 25\% on human preference. On the three novel tasks, we show that AnyV2V also achieves a high success rate. We believe AnyV2V will continue to thrive due to its ability to seamlessly integrate the fast-evolving image editing methods. Such compatibility can help AnyV2V to increase its versatility to cater to diverse user demands.

link: http://arxiv.org/abs/2403.14468v1

ChatGPT Alternative Solutions: Large Language Models Survey

Hanieh Alipour, Nick Pendar, Kohinoor Roy

In recent times, the grandeur of Large Language Models (LLMs) has not only shone in the realm of natural language processing but has also cast its brilliance across a vast array of applications. This remarkable display of LLM capabilities has ignited a surge in research contributions within this domain, spanning a diverse spectrum of topics. These contributions encompass advancements in neural network architecture, context length enhancements, model alignment, training datasets, benchmarking, efficiency improvements, and more. Recent years have witnessed a dynamic synergy between academia and industry, propelling the field of LLM research to new heights. A notable milestone in this journey is the introduction of ChatGPT, a powerful AI chatbot grounded in LLMs, which has garnered widespread societal attention. The evolving technology of LLMs has begun to reshape the landscape of the entire AI community, promising a revolutionary shift in the way we create and employ AI algorithms. Given this swift-paced technical evolution, our survey embarks on a journey to encapsulate the recent strides made in the world of LLMs. Through an exploration of the background, key discoveries, and prevailing methodologies, we offer an up-to-the-minute review of the literature. By examining multiple LLM models, our paper not only

presents a comprehensive overview but also charts a course that identifies existing challenges and points toward potential future research trajectories. This survey furnishes a well-rounded perspective on the current state of generative AI, shedding light on opportunities for further exploration, enhancement, and innovation.

link: http://dx.doi.org/10.5121/csit.2024.140514

Detoxifying Large Language Models via Knowledge Editing

Mengru Wang, Ningyu Zhang, Ziwen Xu, Zekun Xi, Shumin Deng, Yunzhi Yao, Qishen Zhang, Linyi Yang, Jindong Wang, Huajun Chen

This paper investigates using knowledge editing techniques to detoxify Large Language Models (LLMs). We construct a benchmark, SafeEdit, which covers nine unsafe categories with various powerful attack prompts and equips comprehensive metrics for systematic evaluation. We conduct experiments to compare knowledge editing approaches with previous baselines, indicating that knowledge editing has the potential to efficiently detoxify LLMs with limited impact on general performance. Then, we propose a simple yet effective baseline, dubbed Detoxifying with Intraoperative Neural Monitoring (DINM), to diminish the toxicity of LLMs within a few tuning steps via only one instance. We further provide an in-depth analysis of the internal mechanism for various detoxify approaches, demonstrating that previous methods like SFT and DPO may merely suppress the activations of toxic parameters, while DINM mitigates the toxicity of the toxic parameters to a certain extent, making permanent adjustments. We hope that these insights could shed light on future work of developing detoxifying approaches and the underlying knowledge mechanisms of LLMs. Code and benchmark are available at https://github.com/zjunlp/EasyEdit.

link: http://arxiv.org/abs/2403.14472v1

Utilizing the LightGBM Algorithm for Operator User Credit Assessment Research

Shaojie Li, Xinqi Dong, Danqing Ma, Bo Dang, Hengyi Zang, Yulu Gong

Mobile Internet user credit assessment is an important way for communication operators to establish decisions and formulate measures, and it is also a guarantee for operators to obtain expected benefits. However, credit evaluation methods have long been monopolized by financial industries such as banks and credit. As supporters and providers of platform network technology and network resources, communication operators are also builders and maintainers of communication networks. Internet data improves the user's credit evaluation strategy. This paper uses the massive data provided by communication operators to carry out research on the operator's user credit evaluation model based on the fusion LightGBM algorithm. First, for the massive data related to user evaluation provided by operators, key features are extracted by data preprocessing and feature engineering methods, and a multi-dimensional feature set with statistical significance is constructed; then, linear regression, decision tree, LightGBM, and other machine learning algorithms build multiple basic models to find the best basic model; finally, integrates Averaging, Voting, Blending, Stacking and other integrated algorithms to refine multiple fusion models, and finally establish the most suitable fusion model for operator user evaluation.

link: http://arxiv.org/abs/2403.14483v1

HyperGALE: ASD Classification via Hypergraph Gated Attention with Learnable Hyperedges

Mehul Arora, Chirag Shantilal Jain, Lalith Bharadwaj Baru, Kamalaker Dadi, Bapi Raju Surampudi

Autism Spectrum Disorder (ASD) is a neurodevelopmental condition characterized by varied social cognitive challenges and repetitive behavioral patterns. Identifying reliable brain imaging-based biomarkers for ASD has been a persistent challenge due to the spectrum's diverse symptomatology. Existing baselines in the field have made significant strides in this direction, yet there remains room for improvement in both performance and interpretability. We propose \emph{HyperGALE}, which builds upon the hypergraph by incorporating learned hyperedges and gated attention mechanisms. This approach has led to substantial improvements in the model's ability to interpret complex brain graph data, offering deeper insights into ASD biomarker

characterization. Evaluated on the extensive ABIDE II dataset, \emph{HyperGALE} not only improves interpretability but also demonstrates statistically significant enhancements in key performance metrics compared to both previous baselines and the foundational hypergraph model. The advancement \emph{HyperGALE} brings to ASD research highlights the potential of sophisticated graph-based techniques in neurodevelopmental studies. The source code and implementation instructions are available at GitHub:https://github.com/mehular0ra/HyperGALE.

link: http://arxiv.org/abs/2403.14484v1

DesignEdit: Multi-Layered Latent Decomposition and Fusion for Unified & Accurate Image Editing

Yueru Jia, Yuhui Yuan, Aosong Cheng, Chuke Wang, Ji Li, Huizhu Jia, Shanghang Zhang

Recently, how to achieve precise image editing has attracted increasing attention, especially given the remarkable success of text-to-image generation models. To unify various spatial-aware image editing abilities into one framework, we adopt the concept of layers from the design domain to manipulate objects flexibly with various operations. The key insight is to transform the spatial-aware image editing task into a combination of two sub-tasks: multi-layered latent decomposition and multi-layered latent fusion. First, we segment the latent representations of the source images into multiple layers, which include several object layers and one incomplete background layer that necessitates reliable inpainting. To avoid extra tuning, we further explore the inner inpainting ability within the self-attention mechanism. We introduce a key-masking self-attention scheme that can propagate the surrounding context information into the masked region while mitigating its impact on the regions outside the mask. Second, we propose an instruction-guided latent fusion that pastes the multi-layered latent representations onto a canvas latent. We also introduce an artifact suppression scheme in the latent space to enhance the inpainting quality. Due to the inherent modular advantages of such multi-layered representations, we can achieve accurate image editing, and we demonstrate that our approach consistently surpasses the latest spatial editing methods, including Self-Guidance and DiffEditor. Last, we show that our approach is a unified framework that supports various accurate image editing tasks on more than six different editing tasks.

link: http://arxiv.org/abs/2403.14487v1

Physics-Based Causal Reasoning for Safe & Robust Next-Best Action Selection in Robot Manipulation Tasks

Ricardo Cannizzaro, Michael Groom, Jonathan Routley, Robert Osazuwa Ness, Lars Kunze

Safe and efficient object manipulation is a key enabler of many real-world robot applications. However, this is challenging because robot operation must be robust to a range of sensor and actuator uncertainties. In this paper, we present a physics-informed causal-inference-based framework for a robot to probabilistically reason about candidate actions in a block stacking task in a partially observable setting. We integrate a physics-based simulation of the rigid-body system dynamics with a causal Bayesian network (CBN) formulation to define a causal generative probabilistic model of the robot decision-making process. Using simulation-based Monte Carlo experiments, we demonstrate our framework's ability to successfully: (1) predict block tower stability with high accuracy (Pred Acc: 88.6%); and, (2) select an approximate next-best action for the block stacking task, for execution by an integrated robot system, achieving 94.2% task success rate. We also demonstrate our framework's suitability for real-world robot systems by demonstrating successful task executions with a domestic support robot, with perception and manipulation sub-system integration. Hence, we show that by embedding physics-based causal reasoning into robots' decision-making processes, we can make robot task execution safer, more reliable, and more robust to various types of uncertainty.

link: http://arxiv.org/abs/2403.14488v1

Adversary-Robust Graph-Based Learning of WSIs

Saba Heidari Gheshlaghi, Milan Aryal, Nasim Yahyasoltani, Masoud Ganji

Enhancing the robustness of deep learning models against adversarial attacks is crucial, especially in critical domains like healthcare where significant financial interests heighten the risk of such attacks. Whole slide images (WSIs) are high-resolution, digitized versions of tissue samples mounted on glass slides, scanned using sophisticated imaging equipment. The digital analysis of WSIs presents unique challenges due to their gigapixel size and multi-resolution storage format. In this work, we aim at improving the robustness of cancer Gleason grading classification systems against adversarial attacks, addressing challenges at both the image and graph levels. As regards the proposed algorithm, we develop a novel and innovative graph-based model which utilizes GNN to extract features from the graph representation of WSIs. A denoising module, along with a pooling layer is incorporated to manage the impact of adversarial attacks on the WSIs. The process concludes with a transformer module that classifies various grades of prostate cancer based on the processed data. To assess the effectiveness of the proposed method, we conducted a comparative analysis using two scenarios. Initially, we trained and tested the model without the denoiser using WSIs that had not been exposed to any attack. We then introduced a range of attacks at either the image or graph level and processed them through the proposed network. The performance of the model was evaluated in terms of accuracy and kappa scores. The results from this comparison showed a significant improvement in cancer diagnosis accuracy, highlighting the robustness and efficiency of the proposed method in handling adversarial challenges in the context of medical imaging.

link: http://arxiv.org/abs/2403.14489v1