Sat 2024.03.30

INEXA: Interactive and Explainable Process Model Abstraction Through Object-Centric Process Mining

Janik-Vasily Benzin, Gyunam Park, Juergen Mangler, Stefanie Rinderle-Ma

Process events are recorded by multiple information systems at different granularity levels. Based on the resulting event logs, process models are discovered at different granularity levels, as well. Events stored at a fine-grained granularity level, for example, may hinder the discovered process model to be displayed due the high number of resulting model elements. The discovered process model of a real-world manufacturing process, for example, consists of 1,489 model elements and over 2,000 arcs. Existing process model abstraction techniques could help reducing the size of the model, but would disconnect it from the underlying event log. Existing event abstraction techniques do neither support the analysis of mixed granularity levels, nor interactive exploration of a suitable granularity level. To enable the exploration of discovered process models at different granularity levels, we propose INEXA, an interactive, explainable process model abstraction method that keeps the link to the event log. As a starting point, INEXA aggregates large process models to a "displayable" size, e.g., for the manufacturing use case to a process model with 58 model elements. Then, the process analyst can explore granularity levels interactively, while applied abstractions are automatically traced in the event log for explainability.

link: http://arxiv.org/abs/2403.18659v1

InstructBrush: Learning Attention-based Instruction Optimization for Image Editing

Ruoyu Zhao, Qingnan Fan, Fei Kou, Shuai Qin, Hong Gu, Wei Wu, Pengcheng Xu, Mingrui Zhu, Nannan Wang, Xinbo Gao

In recent years, instruction-based image editing methods have garnered significant attention in image editing. However, despite encompassing a wide range of editing priors, these methods are helpless when handling editing tasks that are challenging to accurately describe through language. We propose InstructBrush, an inversion method for instruction-based image editing methods to bridge this gap. It extracts editing effects from exemplar image pairs as editing instructions, which are further applied for image editing. Two key techniques are introduced into InstructBrush, Attention-based Instruction Optimization and Transformation-oriented Instruction Initialization, to address the limitations of the previous method in terms of inversion effects and instruction generalization. To explore the ability of instruction inversion methods to guide image editing in open scenarios, we establish a TransformationOriented Paired Benchmark (TOP-Bench), which contains a rich set of scenes and editing types. The creation of this benchmark paves the way for further exploration of instruction inversion. Quantitatively and qualitatively, our approach achieves superior performance in editing and is more semantically consistent with the target editing effects.

link: http://arxiv.org/abs/2403.18660v1

Neural Network-Based Piecewise Survival Models

Olov Holmer, Erik Frisk, Mattias Krysander

In this paper, a family of neural network-based survival models is presented. The models are specified based on piecewise definitions of the hazard function and the density function on a partitioning of the time; both constant and linear piecewise definitions are presented, resulting in a family of four models. The models can be seen as an extension of the commonly used discrete-time and piecewise exponential models and thereby add flexibility to this set of standard models. Using a simulated dataset the models are shown to perform well compared to the highly expressive, state-of-the-art energy-based model, while only requiring a fraction of the computation time.

link: http://arxiv.org/abs/2403.18664v1

Improving Content Recommendation: Knowledge Graph-Based Semantic Contrastive Learning for Diversity and Cold-Start Users

Yejin Kim, Scott Rome, Kevin Foley, Mayur Nankani, Rimon Melamed, Javier Morales, Abhay Yadav, Maria Peifer, Sardar Hamidian, H. Howie Huang

Addressing the challenges related to data sparsity, cold-start problems, and diversity in recommendation systems is both crucial and demanding. Many current solutions leverage knowledge graphs to tackle these issues by combining both item-based and user-item collaborative signals. A common trend in these approaches focuses on improving ranking performance at the cost of escalating model complexity, reducing diversity, and complicating the task. It is essential to provide recommendations that are both personalized and diverse, rather than solely relying on achieving high rank-based performance, such as Click-through Rate, Recall, etc. In this paper, we propose a hybrid multi-task learning approach, training on user-item and item-item interactions. We apply item-based contrastive learning on descriptive text, sampling positive and negative pairs based on item metadata. Our approach allows the model to better understand the relationships between entities within the knowledge graph by utilizing semantic information from text. It leads to more accurate, relevant, and diverse user recommendations and a benefit that extends even to cold-start users who have few interactions with items. We perform extensive experiments on two widely used datasets to validate the effectiveness of our approach. Our findings demonstrate that jointly training user-item interactions and item-based signals using synopsis text is highly effective. Furthermore, our results provide evidence that item-based contrastive learning enhances the quality of entity embeddings, as indicated by metrics such as uniformity and alignment.

link: http://arxiv.org/abs/2403.18667v1

Aiming for Relevance

Bar Eini Porat, Danny Eytan, Uri Shalit

Vital signs are crucial in intensive care units (ICUs). They are used to track the patient's state and to identify clinically significant changes. Predicting vital sign trajectories is valuable for early detection of adverse events. However, conventional machine learning metrics like RMSE often fail to capture the true clinical relevance of such predictions. We introduce novel vital sign prediction performance metrics that align with clinical contexts, focusing on deviations from clinical norms, overall trends, and trend deviations. These metrics are derived from empirical utility curves obtained in a previous study through interviews with ICU clinicians. We validate the metrics' usefulness using simulated and real clinical datasets (MIMIC and eICU). Furthermore, we employ these metrics as loss functions for neural networks, resulting in models that excel in predicting clinically significant events. This research paves the way for clinically relevant machine learning model evaluation and optimization, promising to improve ICU patient care. 10 pages, 9 figures.

link: http://arxiv.org/abs/2403.18668v1

Fact Checking Beyond Training Set

Payam Karisani, Heng Ji

Evaluating the veracity of everyday claims is time consuming and in some cases requires domain expertise. We empirically demonstrate that the commonly used fact checking pipeline, known as the retriever-reader, suffers from performance deterioration when it is trained on the labeled data from one domain and used in another domain. Afterwards, we delve into each component of the pipeline and propose novel algorithms to address this problem. We propose an adversarial algorithm to make the retriever component robust against distribution shift. Our core idea is to initially train a bi-encoder on the labeled source data, and then, to adversarially train two separate document and claim encoders using unlabeled target data. We then focus on the reader component and propose to train it such that it is insensitive towards the order of claims and evidence documents. Our empirical evaluations support the hypothesis that such a reader shows a higher robustness against distribution shift. To our knowledge, there is no publicly available multi-topic fact checking dataset. Thus, we propose a simple automatic method to re-purpose two well-known fact checking datasets. We then construct eight fact checking scenarios from these datasets, and compare our model to a set of strong baseline models, including recent domain adaptation models

that use GPT4 for generating synthetic data.

link: http://arxiv.org/abs/2403.18671v1

Deep Learning for Robust and Explainable Models in Computer Vision

Mohammadreza Amirian

Recent breakthroughs in machine and deep learning (ML and DL) research have provided excellent tools for leveraging enormous amounts of data and optimizing huge models with millions of parameters to obtain accurate networks for image processing. These developments open up tremendous opportunities for using artificial intelligence (AI) in the automation and human assisted AI industry. However, as more and more models are deployed and used in practice, many challenges have emerged. This thesis presents various approaches that address robustness and explainability challenges for using ML and DL in practice. Robustness and reliability are the critical components of any model before certification and deployment in practice. Deep convolutional neural networks (CNNs) exhibit vulnerability to transformations of their inputs, such as rotation and scaling, or intentional manipulations as described in the adversarial attack literature. In addition, building trust in AI-based models requires a better understanding of current models and developing methods that are more explainable and interpretable a priori. This thesis presents developments in computer vision models' robustness and explainability. Furthermore, this thesis offers an example of using vision models' feature response visualization (models' interpretations) to improve robustness despite interpretability and robustness being seemingly unrelated in the related research. Besides methodological developments for robust and explainable vision models, a key message of this thesis is introducing model interpretation techniques as a tool for understanding vision models and improving their design and robustness. In addition to the theoretical developments, this thesis demonstrates several applications of ML and DL in different contexts, such as medical imaging and affective computing.

link: http://dx.doi.org/10.18725/OPARU-51464

NL-ITI: Optimizing Probing and Intervention for Improvement of ITI Method

Jakub Hoscilowicz, Adam Wiacek, Jan Chojnacki, Adam Cieslak, Leszek Michon, Vitalii Urbanevych, Artur Janicki

Large Language Models (LLM) are prone to returning false information. It constitutes one of major challenges in the AI field. In our work, we explore paradigm introduced by Inference-Time-Intervention (ITI). In first stage, it identifies attention heads, which contain the highest amount of desired type of knowledge (e.g., truthful). Afterwards, during inference, LLM activations are shifted for chosen subset of attention heads. We further improved the ITI framework by introducing a nonlinear probing and multi-token intervention - Non-Linear ITI (NL-ITI). NL-ITI is tested on diverse multiple-choice benchmarks, including TruthfulQA, on which we report around 14% MC1 metric improvement with respect to the baseline ITI results. NL-ITI achieves also encouraging results on other testsets - on Business Ethics subdomain of MMLU, around 18% MC1 improvement over baseline LLaMA2-7B. Additionally, NL-ITI performs better while being less invasive in the behavior of LLM at the same time (as measured by Kullback-Leibler divergence).

link: http://arxiv.org/abs/2403.18680v1

TransFusion: Contrastive Learning with Transformers

Huanran Li, Daniel Pimentel-Alarcón

This paper proposes a novel framework, TransFusion, designed to make the process of contrastive learning more analytical and explainable. TransFusion consists of attention blocks whose softmax being replaced by ReLU, and its final block's weighted-sum operation is truncated to leave the adjacency matrix as the output. The model is trained by minimizing the Jensen-Shannon Divergence between its output and the target affinity matrix, which indicates whether each pair of samples belongs to the same or different classes. The main contribution of TransFusion lies in defining a theoretical limit for answering two fundamental questions in the field: the maximum level of data augmentation and the minimum batch size required for effective contrastive learning.

Furthermore, experimental results indicate that TransFusion successfully extracts features that isolate clusters from complex real-world data, leading to improved classification accuracy in downstream tasks.

link: http://arxiv.org/abs/2403.18681v1

JumpBackHash: Say Goodbye to the Modulo Operation to Distribute Keys Uniformly to Buckets

Otmar Ertl

The distribution of keys to a given number of buckets is a fundamental task in distributed data processing and storage. A simple, fast, and therefore popular approach is to map the hash values of keys to buckets based on the remainder after dividing by the number of buckets. Unfortunately, these mappings are not stable when the number of buckets changes, which can lead to severe spikes in system resource utilization, such as network or database requests. Consistent hash algorithms can minimize remappings, but are either significantly slower than the modulo-based approach, require floating-point arithmetic, or are based on a family of hash functions rarely available in standard libraries. This paper introduces JumpBackHash, which uses only integer arithmetic and a standard pseudorandom generator. Due to its speed and simple implementation, it can safely replace the modulo-based approach to improve assignment and system stability. A production-ready Java implementation of JumpBackHash has been released as part of the Hash4j open source library.

link: http://arxiv.org/abs/2403.18682v1

Scaling Laws For Dense Retrieval

Yan Fang, Jingtao Zhan, Qingyao Ai, Jiaxin Mao, Weihang Su, Jia Chen, Yiqun Liu

Scaling up neural models has yielded significant advancements in a wide array of tasks, particularly in language generation. Previous studies have found that the performance of neural models frequently adheres to predictable scaling laws, correlated with factors such as training set size and model size. This insight is invaluable, especially as large-scale experiments grow increasingly resource-intensive. Yet, such scaling law has not been fully explored in dense retrieval due to the discrete nature of retrieval metrics and complex relationships between training data and model sizes in retrieval tasks. In this study, we investigate whether the performance of dense retrieval models follows the scaling law as other neural models. We propose to use contrastive log-likelihood as the evaluation metric and conduct extensive experiments with dense retrieval models implemented with different numbers of parameters and trained with different amounts of annotated data. Results indicate that, under our settings, the performance of dense retrieval models follows a precise power-law scaling related to the model size and the number of annotations. Additionally, we examine scaling with prevalent data augmentation methods to assess the impact of annotation quality, and apply the scaling law to find the best resource allocation strategy under a budget constraint. We believe that these insights will significantly contribute to understanding the scaling effect of dense retrieval models and offer meaningful guidance for future research endeavors.

link: http://arxiv.org/abs/2403.18684v1

Representatividad Muestral en la Incertidumbre Simétrica Multivariada para la Selección de Atributos

Gustavo Sosa-Cabrera

In this work, we analyze the behavior of the multivariate symmetric uncertainty (MSU) measure through the use of statistical simulation techniques under various mixes of informative and non-informative randomly generated features. Experiments show how the number of attributes, their cardinalities, and the sample size affect the MSU. In this thesis, through observation of results, it is proposed an heuristic condition that preserves good quality in the MSU under different combinations of these three factors, providing a new useful criterion to help drive the process of dimension reduction. -- En el presente trabajo hemos analizado el comportamiento de una versi\'on multivariada de la incertidumbre sim\'etrica a trav\'es de t\'ecnicas de simulaci\'on estad\'isticas

sobre varias combinaciones de atributos informativos y no-informativos generados de forma aleatoria. Los experimentos muestran como el n\'umero de atributos, sus cardinalidades y el tama\~no muestral afectan al MSU como medida. En esta tesis, mediante la observaci\'on de resultados hemos propuesto una condici\'on que preserva una buena calidad en el MSU bajo diferentes combinaciones de los tres factores mencionados, lo cual provee un nuevo y valioso criterio para llevar a cabo el proceso de reducci\'on de dimensionalidad.

link: http://arxiv.org/abs/2403.18685v1

InceptionTime vs. Wavelet -- A comparison for time series classification

Daniel Klenkert, Daniel Schaeffer, Julian Stauch

Neural networks were used to classify infrasound data. Two different approaches were compared. One based on the direct classification of time series data, using a custom implementation of the InceptionTime network. For the other approach, we generated 2D images of the wavelet transformation of the signals, which were subsequently classified using a ResNet implementation. Choosing appropriate hyperparameter settings, both achieve a classification accuracy of above 90 %, with the direct approach reaching 95.2 %.

link: http://arxiv.org/abs/2403.18687v1

Annolid: Annotate, Segment, and Track Anything You Need

Chen Yang, Thomas A. Cleland

Annolid is a deep learning-based software package designed for the segmentation, labeling, and tracking of research targets within video files, focusing primarily on animal behavior analysis. Based on state-of-the-art instance segmentation methods, Annolid now harnesses the Cutie video object segmentation model to achieve resilient, markerless tracking of multiple animals from single annotated frames, even in environments in which they may be partially or entirely concealed by environmental features or by one another. Our integration of Segment Anything and Grounding-DINO strategies additionally enables the automatic masking and segmentation of recognizable animals and objects by text command, removing the need for manual annotation. Annolid's comprehensive approach to object segmentation flexibly accommodates a broad spectrum of behavior analysis applications, enabling the classification of diverse behavioral states such as freezing, digging, pup huddling, and social interactions in addition to the tracking of animals and their body parts.

link: http://arxiv.org/abs/2403.18690v1

The Invalsi Benchmark: measuring Language Models Mathematical and Language understanding in Italian

Andrea Esuli, Giovanni Puccetti

While Italian is by all metrics a high resource language, currently, there are isn't a Language Model pre-trained exclusively in this language. This results in a lower number of available benchmarks to evaluate the performance of language models in Italian. This work presents two new benchmarks to evaluate the models performance on mathematical understanding and language understanding in Italian. These benchmarks are based on real tests that are undertaken by students of age between 11 and 18 within the Italian school system and have therefore been validated by several experts in didactics and pedagogy. To validate this dataset we evaluate the performance of 9 language models that are the best performing when writing in Italian, including our own fine-tuned models. We show that this is a challenging benchmark where current language models are bound by 60\% accuracy. We believe that the release of this dataset paves the way for improving future models mathematical and language understanding in Italian.

link: http://arxiv.org/abs/2403.18697v1