Sun 2024.04.21

A Time-Inhomogeneous Markov Model for Resource Availability under Sparse Observations

Lukas Rottkamp, Matthias Schubert

Accurate spatio-temporal information about the current situation is crucial for smart city applications such as modern routing algorithms. Often, this information describes the state of stationary resources, e.g. the availability of parking bays, charging stations or the amount of people waiting for a vehicle to pick them up near a given location. To exploit this kind of information, predicting future states of the monitored resources is often mandatory because a resource might change its state within the time until it is needed. To train an accurate predictive model, it is often not possible to obtain a continuous time series on the state of the resource. For example, the information might be collected from traveling agents visiting the resource with an irregular frequency. Thus, it is necessary to develop methods which work on sparse observations for training and prediction. In this paper, we propose time-inhomogeneous discrete Markov models to allow accurate prediction even when the frequency of observation is very rare. Our new model is able to blend recent observations with historic data and also provide useful probabilistic estimates for future states. Since resources availability in a city is typically time-dependent, our Markov model is time-inhomogeneous and cyclic within a predefined time interval. To train our model, we propose a modified Baum-Welch algorithm. Evaluations on real-world datasets of parking bay availability show that our new method indeed yields good results compared to methods being trained on complete data and non-cyclic variants.

link: http://dx.doi.org/10.1145/3274895.3274945

Introducing v0.5 of the AI Safety Benchmark from MLCommons

Bertie Vidgen, Adarsh Agrawal, Ahmed M. Ahmed, Victor Akinwande, Namir Al-Nuaimi, Najla Alfaraj, Elie Alhajjar, Lora Aroyo, Trupti Bavalatti, Borhane Blili-Hamelin, Kurt Bollacker, Rishi Bomassani, Marisa Ferrara Boston, Siméon Campos, Kal Chakra, Canyu Chen, Cody Coleman, Zacharie Delpierre Coudert, Leon Derczynski, Debojyoti Dutta, Ian Eisenberg, James Ezick, Heather Frase, Brian Fuller, Ram Gandikota, Agasthya Gangavarapu, Ananya Gangavarapu, James Gealy, Rajat Ghosh, James Goel, Usman Gohar, Sujata Goswami, Scott A. Hale, Wiebke Hutiri, Joseph Marvin Imperial, Surgan Jandial, Nick Judd, Felix Juefei-Xu, Foutse Khomh, Bhavya Kailkhura, Hannah Rose Kirk, Kevin Klyman, Chris Knotz, Michael Kuchnik, Shachi H. Kumar, Chris Lengerich, Bo Li, Zeyi Liao, Eileen Peters Long, Victor Lu, Yifan Mai, Priyanka Mary Mammen, Kelvin Manyeki, Sean McGregor, Virendra Mehta, Shafee Mohammed, Emanuel Moss, Lama Nachman, Dinesh Jinenhally Naganna, Amin Nikanjam, Besmira Nushi, Luis Oala, Iftach Orr, Alicia Parrish, Cigdem Patlak, William Pietri, Forough Poursabzi-Sangdeh, Eleonora Presani, Fabrizio Puletti, Paul Röttger, Saurav Sahay, Tim Santos, Nino Scherrer, Alice Schoenauer Sebag, Patrick Schramowski, Abolfazl Shahbazi, Vin Sharma, Xudong Shen, Vamsi Sistla, Leonard Tang, Davide Testuggine, Vithursan Thangarasa, Elizabeth Anne Watkins, Rebecca Weiss, Chris Welty, Tyler Wilbers, Adina Williams, Carole-Jean Wu, Poonam Yadav, Xianjun Yang, Yi Zeng, Wenhui Zhang, Fedor Zhdanov, Jiacheng Zhu, Percy Liang, Peter Mattson, Joaquin Vanschoren

This paper introduces v0.5 of the AI Safety Benchmark, which has been created by the MLCommons AI Safety Working Group. The AI Safety Benchmark has been designed to assess the safety risks of AI systems that use chat-tuned language models. We introduce a principled approach to specifying and constructing the benchmark, which for v0.5 covers only a single use case (an adult chatting to a general-purpose assistant in English), and a limited set of personas (i.e., typical users, malicious users, and vulnerable users). We created a new taxonomy of 13 hazard categories, of which 7 have tests in the v0.5 benchmark. We plan to release version 1.0 of the AI Safety Benchmark by the end of 2024. The v1.0 benchmark will provide meaningful insights into the safety of AI systems. However, the v0.5 benchmark should not be used to assess the safety of AI systems. We have sought to fully document the limitations, flaws, and challenges of v0.5. This release of v0.5 of the AI Safety Benchmark includes (1) a principled approach to specifying and constructing the benchmark, which comprises use cases, types of systems under test (SUTs), language and context, personas, tests, and test items; (2) a taxonomy of 13 hazard categories with definitions and subcategories; (3) tests for seven of the hazard categories, each comprising a unique set of test items, i.e., prompts. There are 43,090 test items in total, which we created with templates; (4) a grading system for AI systems against the benchmark; (5) an openly available platform, and downloadable tool, called ModelBench that can be used to evaluate the safety of AI systems on the benchmark; (6) an example evaluation report which benchmarks the performance of over a dozen openly available chat-tuned language models; (7) a test specification for the benchmark.

link: http://arxiv.org/abs/2404.12241v1

CMNEE: A Large-Scale Document-Level Event Extraction Dataset based on Open-Source Chinese Military News

Mengna Zhu, Zijie Xu, Kaisheng Zeng, Kaiming Xiao, Mao Wang, Wenjun Ke, Hongbin Huang

Extracting structured event knowledge, including event triggers and corresponding arguments, from military texts is fundamental to many applications, such as intelligence analysis and decision assistance. However, event extraction in the military field faces the data scarcity problem, which impedes the research of event extraction models in this domain. To alleviate this problem, we propose CMNEE, a large-scale, document-level open-source Chinese Military News Event Extraction dataset. It contains 17,000 documents and 29,223 events, which are all manually annotated based on a pre-defined schema for the military domain including 8 event types and 11 argument role types. We designed a two-stage, multi-turns annotation strategy to ensure the quality of CMNEE and reproduced several state-of-the-art event extraction models with a systematic evaluation. The experimental results on CMNEE fall shorter than those on other domain datasets obviously, which demonstrates that event extraction for military domain poses unique challenges and requires further research efforts. Our code and data can be obtained from https://github.com/Mzzzhu/CMNEE.

link: http://arxiv.org/abs/2404.12242v1

Blind Localization and Clustering of Anomalies in Textures

Andrei-Timotei Ardelean, Tim Weyrich

Anomaly detection and localization in images is a growing field in computer vision. In this area, a seemingly understudied problem is anomaly clustering, i.e., identifying and grouping different types of anomalies in a fully unsupervised manner. In this work, we propose a novel method for clustering anomalies in largely stationary images (textures) in a blind setting. That is, the input consists of normal and anomalous images without distinction and without labels. What contributes to the difficulty of the task is that anomalous regions are often small and may present only subtle changes in appearance, which can be easily overshadowed by the genuine variance in the texture. Moreover, each anomaly type may have a complex appearance distribution. We introduce a novel scheme for solving this task using a combination of blind anomaly localization and contrastive learning. By identifying the anomalous regions with high fidelity, we can restrict our focus to those regions of interest; then, contrastive learning is employed to increase the separability of different anomaly types and reduce the intra-class variation. Our experiments show that the proposed solution yields significantly better results compared to prior work, setting a new state of the art. Project page: https://reality.tf.fau.de/pub/ardelean2024blind.html.

link: http://arxiv.org/abs/2404.12246v1

Dynamic Modality and View Selection for Multimodal Emotion Recognition with Missing Modalities

Luciana Trinkaus Menon, Luiz Carlos Ribeiro Neduziak, Jean Paul Barddal, Alessandro Lameiras Koerich, Alceu de Souza Britto Jr

The study of human emotions, traditionally a cornerstone in fields like psychology and neuroscience, has been profoundly impacted by the advent of artificial intelligence (AI). Multiple channels, such as speech (voice) and facial expressions (image), are crucial in understanding human emotions. However, AI's journey in multimodal emotion recognition (MER) is marked by substantial technical challenges. One significant hurdle is how AI models manage the absence of a particular modality - a frequent occurrence in real-world situations. This study's central focus is assessing the performance and resilience of two strategies when confronted with the lack of one modality: a novel multimodal dynamic modality and view selection and a cross-attention mechanism. Results on the RECOLA dataset show that dynamic selection-based methods are a promising approach for MER. In the missing modalities scenarios, all dynamic selection-based methods outperformed the baseline. The study concludes by emphasizing the intricate interplay between audio and video modalities in emotion prediction, showcasing the adaptability of dynamic selection methods in handling missing modalities.

link: http://arxiv.org/abs/2404.12251v1

Deep Gaussian mixture model for unsupervised image segmentation

Matthias Schwab, Agnes Mayr, Markus Haltmeier

The recent emergence of deep learning has led to a great deal of work on designing supervised deep semantic segmentation algorithms. As in many tasks sufficient pixel-level labels are very difficult to obtain, we propose a method which combines a Gaussian mixture model (GMM) with unsupervised deep learning techniques. In the standard GMM the pixel values with each sub-region are modelled by a Gaussian distribution. In order to identify the different regions, the parameter vector that minimizes the negative log-likelihood (NLL) function regarding the GMM has to be approximated. For this task, usually iterative optimization methods such as the expectation-maximization (EM) algorithm are used. In this paper, we propose to estimate these parameters directly from the image using a convolutional neural network (CNN). We thus change the iterative procedure in the EM algorithm replacing the expectation-step by a gradient-step with regard to the networks parameters. This means that the network is trained to minimize the NLL function of the GMM which comes with at least two advantages. As once trained, the network is able to predict label probabilities very quickly compared with time consuming iterative optimization methods. Secondly, due to the deep image prior our method is able to partially overcome one of the main disadvantages of GMM, which is not taking into account correlation between neighboring pixels, as it assumes independence between them. We demonstrate the advantages of our method in various experiments on the example of myocardial infarct segmentation on multi-sequence MRI images.

link: http://arxiv.org/abs/2404.12252v1

Toward Self-Improvement of LLMs via Imagination, Searching, and Criticizing

Ye Tian, Baolin Peng, Linfeng Song, Lifeng Jin, Dian Yu, Haitao Mi, Dong Yu

Despite the impressive capabilities of Large Language Models (LLMs) on various tasks, they still struggle with scenarios that involves complex reasoning and planning. Recent work proposed advanced prompting techniques and the necessity of fine-tuning with high-quality data to augment LLMs' reasoning abilities. However, these approaches are inherently constrained by data availability and quality. In light of this, self-correction and self-learning emerge as viable solutions, employing strategies that allow LLMs to refine their outputs and learn from self-assessed rewards. Yet, the efficacy of LLMs in self-refining its response, particularly in complex reasoning and planning task, remains dubious. In this paper, we introduce AlphaLLM for the self-improvements of LLMs, which integrates Monte Carlo Tree Search (MCTS) with LLMs to establish a self-improving loop, thereby enhancing the capabilities of LLMs without additional annotations. Drawing inspiration from the success of AlphaGo, AlphaLLM addresses the unique challenges of combining MCTS with LLM for self-improvement, including data scarcity, the vastness search spaces of language tasks, and the subjective nature of feedback in language tasks. AlphaLLM is comprised of prompt synthesis component, an efficient MCTS approach tailored for language tasks, and a trio of critic models for precise feedback. Our experimental results in mathematical reasoning tasks demonstrate that AlphaLLM significantly enhances the performance of LLMs without additional annotations, showing the potential for self-improvement in LLMs.

link: http://arxiv.org/abs/2404.12253v1

An Online Spatial-Temporal Graph Trajectory Planner for Autonomous Vehicles

Jilan Samiuddin, Benoit Boulet, Di Wu

The autonomous driving industry is expected to grow by over 20 times in the coming decade and, thus, motivate researchers to delve into it. The primary focus of their research is to ensure safety, comfort, and efficiency. An autonomous vehicle has several modules responsible for one or more of the aforementioned items. Among these modules, the trajectory planner plays a pivotal role in the safety of the vehicle and the comfort of its passengers. The module is also responsible for respecting kinematic constraints and any applicable road constraints. In this paper, a novel online spatial-temporal graph trajectory planner is introduced to generate safe and comfortable trajectories. First, a spatial-temporal graph is constructed using the autonomous vehicle, its surrounding vehicles, and virtual nodes along the road with respect to the vehicle itself. Next, the graph is forwarded into a sequential network to obtain the desired states. To support the planner, a simple behavioral layer is also presented that determines kinematic constraints for the planner. Furthermore, a novel potential function is also proposed to train the network. Finally, the proposed planner is tested on three different complex driving tasks, and the performance is compared with two frequently used methods. The results show that the proposed planner generates safe and feasible trajectories while achieving similar or longer distances in the forward direction and comparable comfort ride.

link: http://dx.doi.org/10.1109/TIV.2024.3389640

Food Portion Estimation via 3D Object Scaling

Gautham Vinod, Jiangpeng He, Zeman Shao, Fengqing Zhu

Image-based methods to analyze food images have alleviated the user burden and biases associated with traditional methods. However, accurate portion estimation remains a major challenge due to the loss of 3D information in the 2D representation of foods captured by smartphone cameras or wearable devices. In this paper, we propose a new framework to estimate both food volume and energy from 2D images by leveraging the power of 3D food models and physical reference in the eating scene. Our method estimates the pose of the camera and the food object in the input image and recreates the eating occasion by rendering an image of a 3D model of the food with the estimated poses. We also introduce a new dataset, SimpleFood45, which contains 2D images of 45 food items and associated annotations including food volume, weight, and energy. Our method achieves an average error of 31.10 kCal (17.67%) on this dataset, outperforming existing portion estimation methods.

link: http://arxiv.org/abs/2404.12257v1

DeepLocalization: Using change point detection for Temporal Action Localization

Mohammed Shaiqur Rahman, Ibne Farabi Shihab, Lynna Chu, Anuj Sharma

In this study, we introduce DeepLocalization, an innovative framework devised for the real-time localization of actions tailored explicitly for monitoring driver behavior. Utilizing the power of advanced deep learning methodologies, our objective is to tackle the critical issue of distracted driving-a significant factor contributing to road accidents. Our strategy employs a dual approach: leveraging Graph-Based Change-Point Detection for pinpointing actions in time alongside a Video Large Language Model (Video-LLM) for precisely categorizing activities. Through careful prompt engineering, we customize the Video-LLM to adeptly handle driving activities' nuances, ensuring its classification efficacy even with sparse data. Engineered to be lightweight, our framework is optimized for consumer-grade GPUs, making it vastly applicable in practical scenarios. We subjected our method to rigorous testing on the SynDD2 dataset, a complex benchmark for distracted driving behaviors, where it demonstrated commendable performance-achieving 57.5% accuracy in event classification and 51% in event detection. These outcomes underscore the substantial promise of DeepLocalization in accurately identifying diverse driver behaviors and their temporal occurrences, all within the bounds of limited computational resources.

link: http://arxiv.org/abs/2404.12258v1

Concept Induction: Analyzing Unstructured Text with High-Level Concepts Using LLooM

Michelle S. Lam, Janice Teoh, James Landay, Jeffrey Heer, Michael S. Bernstein

Data analysts have long sought to turn unstructured text data into meaningful concepts. Though common, topic modeling and clustering focus on lower-level keywords and require significant interpretative work. We introduce concept induction, a computational process that instead produces high-level concepts, defined by explicit inclusion criteria, from unstructured text. For a dataset of toxic online comments, where a state-of-the-art BERTopic model outputs "women, power, female," concept induction produces high-level concepts such as "Criticism of traditional gender roles" and "Dismissal of women's concerns." We present LLooM, a concept induction algorithm that leverages large language models to iteratively synthesize sampled text and propose human-interpretable concepts of increasing generality. We then instantiate LLooM in a mixed-initiative text analysis tool, enabling analysts to shift their attention from interpreting topics to engaging in theory-driven analysis. Through technical evaluations and four analysis scenarios ranging from literature review to content moderation, we find that LLooM's concepts improve upon the prior art of topic models in terms of quality and data coverage. In expert case studies, LLooM helped researchers to uncover new insights even from familiar datasets, for example by suggesting a previously unnoticed concept of attacks on out-party stances in a political social media dataset.

link: http://dx.doi.org/10.1145/3613904.3642830

Alleviating Catastrophic Forgetting in Facial Expression Recognition with Emotion-Centered Models

Israel A. Laurensi, Alceu de Souza Britto Jr., Jean Paul Barddal, Alessandro Lameiras Koerich

Facial expression recognition is a pivotal component in machine learning, facilitating various applications. However, convolutional neural networks (CNNs) are often plagued by catastrophic forgetting, impeding their adaptability. The proposed method, emotion-centered generative replay (ECgr), tackles this challenge by integrating synthetic images from generative adversarial networks. Moreover, ECgr incorporates a quality assurance algorithm to ensure the fidelity of generated images. This dual approach enables CNNs to retain past knowledge while learning new tasks, enhancing their performance in emotion recognition. The experimental results on four diverse facial expression datasets demonstrate that incorporating images generated by our pseudo-rehearsal method enhances training on the targeted dataset and the source dataset while making the CNN retain previously learned knowledge.

link: http://arxiv.org/abs/2404.12260v1

Physics-integrated generative modeling using attentive planar normalizing flow based variational autoencoder

Sheikh Waqas Akhtar

Physics-integrated generative modeling is a class of hybrid or grey-box modeling in which we augment the the data-driven model with the physics knowledge governing the data distribution. The use of physics knowledge allows the generative model to produce output in a controlled way, so that the output, by construction, complies with the physical laws. It imparts improved generalization ability to extrapolate beyond the training distribution as well as improved interpretability because the model is partly grounded in firm domain knowledge. In this work, we aim to improve the fidelity of reconstruction and robustness to noise in the physics integrated generative model. To this end, we use variational-autoencoder as a generative model. To improve the reconstruction results of the decoder, we propose to learn the latent posterior distribution of both the physics as well as the trainable data-driven components using planar normalizng flow. Normalizng flow based posterior distribution harnesses the inherent dynamical structure of the data distribution, hence the learned model gets closer to the true underlying data distribution. To improve the robustness of generative model against noise injected in the model, we propose a modification in the encoder part of the normalizing flow based VAE. We designed the encoder to incorporate scaled dot product attention based contextual information in the noisy latent vector which will mitigate the adverse effect of noise in the latent vector and make the model more robust. We empirically evaluated our models on human locomotion dataset [33] and the results validate the efficacy of our proposed models in terms of improvement in reconstruction quality as well as robustness against noise injected in the model.

link: http://arxiv.org/abs/2404.12267v1

How Population Diversity Influences the Efficiency of Crossover

Sacha Cerf, Johannes Lengler

Our theoretical understanding of crossover is limited by our ability to analyze how population diversity evolves. In this study, we provide one of the first rigorous analyses of population diversity and optimization time in a setting where large diversity and large population sizes are required to speed up progress. We give a formal and general criterion which amount of diversity is necessary and sufficient to speed up the $(\mu+1)$ Genetic Algorithm on LeadingOnes. We show that the naturally evolving diversity falls short of giving a substantial speed-up for any $\mu=O(\sqrt{n}/\log^2 n)$. On the other hand, we show that even for $\mu=2$, if we simply break ties in favor of diversity then this increases diversity so much that optimization is accelerated by a constant factor.

link: http://arxiv.org/abs/2404.12268v1

Who Validates the Validators? Aligning LLM-Assisted Evaluation of LLM Outputs with Human Preferences

Shreya Shankar, J. D. Zamfirescu-Pereira, Björn Hartmann, Aditya G. Parameswaran, Ian Arawjo

Due to the cumbersome nature of human evaluation and limitations of code-based evaluation, Large Language Models (LLMs) are increasingly being used to assist humans in evaluating LLM outputs. Yet LLM-generated evaluators simply inherit all the problems of the LLMs they evaluate, requiring further human validation. We present a mixed-initiative approach to ``validate the validators'' -- aligning LLM-generated evaluation functions (be it prompts or code) with human requirements. Our interface, EvalGen, provides automated assistance to users in generating evaluation criteria and implementing assertions. While generating candidate implementations (Python functions, LLM grader prompts), EvalGen asks humans to grade a subset of LLM outputs; this feedback is used to select implementations that better align with user grades. A qualitative study finds overall support for EvalGen but underscores the subjectivity and iterative process of alignment. In particular, we identify a phenomenon we dub \emph{criteria drift}: users need criteria to grade outputs, but grading outputs helps users define criteria. What is more, some criteria appears \emph{dependent} on the specific LLM outputs observed (rather than independent criteria that can be defined \emph{a priori}), raising serious questions for approaches that assume the independence of evaluation from observation of model outputs. We present our interface and implementation details, a comparison of our algorithm with a baseline approach, and implications for the design of future LLM evaluation assistants.

link: http://arxiv.org/abs/2404.12272v1