Tue 2024.05.07

TV White Space and LTE Network Optimization towards Energy Efficiency in Suburban and Rural Scenarios

Rodney Martinez Alonso, David Plets, Margot Deruyck, Luc Martens, Wout Joseph

The radio spectrum is a limited resource. Demand for wireless communication services is increasing exponentially, stressing the availability of radio spectrum to accommodate new services. TV White Space (TVWS) technologies allow a dynamic usage of the spectrum. These technologies provide wireless connectivity, in the channels of the Very High Frequency (VHF) and Ultra High Frequency (UHF) television broadcasting bands. In this paper, we investigate and compare the coverage range, network capacity, and network energy efficiency for TVWS technologies and LTE. We consider Ghent, Belgium and Boyeros, Havana, Cuba to evaluate a realistic outdoor suburban and rural area, respectively. The comparison shows that TVWS networks have an energy efficiency 9-12 times higher than LTE networks.

link: http://dx.doi.org/10.1109/TBC.2017.2731043

Multi-objective Optimization of Cognitive Radio Networks

Rodney Martinez Alonso, David Plets, Margot Deruyck, Luc Martens, Glauco Guillen Nieto, Wout Joseph

New generation networks, based on Cognitive Radio technology, allow dynamic allocation of the spectrum, alleviating spectrum scarcity. These networks also have a resilient potential for dynamic operation for energy saving. In this paper, we present a novel wireless network optimization algorithm for cognitive radio networks based on a cloud sharing-decision mechanism. Three Key Performance Indicators (KPIs) were optimized: spectrum usage, power consumption, and exposure of human beings. For a realistic suburban scenario in Ghent city, Belgium, we determine the optimality among the KPIs. Compared to a traditional Cognitive Radio network design, our optimization algorithm for the cloud-based architecture reduced the network power consumption by 27.5%, the average global exposure by 34.3%, and spectrum usage by 34.5% at the same time. Even for the worst optimization case, our solution performs better than the traditional architecture by 4.8% in terms of network power consumption, 7.3% in terms of spectrum usage and 4.3% in terms of global exposure.

link: http://dx.doi.org/10.1016/j.comnet.2020.107651

Improved All-Pairs Approximate Shortest Paths in Congested Clique

Hong Duc Bui, Shashwat Chandra, Yi-Jun Chang, Michal Dory, Dean Leitersdorf

In this paper, we present new algorithms for approximating All-Pairs Shortest Paths (APSP) in the Congested Clique model. We present randomized algorithms for weighted undirected graphs. Our first contribution is an $O(1)$-approximate APSP algorithm taking just $O(\log \log \log n)$ rounds. Prior to our work, the fastest algorithms that give an $O(1)$-approximation for APSP take $\operatorname{poly}(\log{n})$ rounds in weighted undirected graphs, and $\operatorname{poly}(\log \log n)$ rounds in unweighted undirected graphs. If we terminate the execution of the algorithm early, we obtain an $O(t)$-round algorithm that yields an $O \big((\log n)^{1/2^t} \big) $ distance approximation for a parameter t. The trade-off between t and the approximation quality provides flexibility for different scenarios, allowing the algorithm to adapt to specific requirements. In particular, we can get an $O \big((\log n)^{1/2^t} \big) $-approximation for any constant t in $O(1)$-rounds. Such result was previously known only for the special case that $t=0$. A key ingredient in our algorithm is a lemma that allows to improve an $O(a)$-approximation for APSP to an $O(\sqrt{a})$-approximation for APSP in $O(1)$ rounds. To prove the lemma, we develop several new tools, including $O(1)$-round algorithms for computing the k closest nodes, a certain type of hopset, and skeleton graphs.

link: http://arxiv.org/abs/2405.02695v1

Stable Diffusion Dataset Generation for Downstream Classification Tasks

Eugenio Lomurno, Matteo D'Oria, Matteo Matteucci

Recent advances in generative artificial intelligence have enabled the creation of high-quality synthetic data that closely mimics real-world data. This paper explores the adaptation of the Stable Diffusion 2.0 model for generating synthetic datasets, using Transfer Learning, Fine-Tuning and generation parameter optimisation techniques to improve the utility of the dataset for downstream classification tasks. We present a class-conditional version of the model that exploits a Class-Encoder and optimisation of key generation parameters. Our methodology led to synthetic datasets that, in a third of cases, produced models that outperformed those trained on real datasets.

link: http://arxiv.org/abs/2405.02698v1

Towards a Scalable Identification of Novel Modes in Generative Models

Jingwei Zhang, Mohammad Jalali, Cheuk Ting Li, Farzan Farnia

An interpretable comparison of generative models requires the identification of sample types produced more frequently by each of the involved models. While several quantitative scores have been proposed in the literature to rank different generative models, such score-based evaluations do not reveal the nuanced differences between the generative models in capturing various sample types. In this work, we propose a method called Fourier-based Identification of Novel Clusters (FINC) to identify modes produced by a generative model with a higher frequency in comparison to a reference distribution. FINC provides a scalable stochastic algorithm based on random Fourier features to estimate the eigenspace of kernel covariance matrices of two generative models and utilize the principal eigendirections to detect the sample types present more dominantly in each model. We demonstrate the application of the FINC method to standard computer vision datasets and generative model frameworks. Our numerical results suggest the scalability and efficiency of the developed Fourier-based method in highlighting the sample types captured with different frequencies by widely-used generative models.

link: http://arxiv.org/abs/2405.02700v1

Enhancing News Summarization with ELearnFit through Efficient In-Context Learning and Efficient Fine-Tuning

Che Guan, Andrew Chin, Puya Vahabi

With the deluge of information delivered by the daily news cycle, there is a growing need to effectively and efficiently summarize news feeds for quick consumption. We leverage large language models (LLMs), with their advanced learning and generative abilities as compared to conventional language models, to generate concise and coherent summaries for news articles from the XSum dataset. Our paper focuses on two key aspects of LLMs: Efficient in-context Learning (ELearn) and Parameter Efficient Fine-tuning (EFit). Under ELearn, we find that increasing the number of shots in prompts and utilizing simple templates generally improve the quality of summaries. We also find that utilizing relevant examples in few-shot learning for ELearn does not improve model performance. In addition, we studied EFit using different methods and demonstrate that fine-tuning the first layer of LLMs produces better outcomes as compared to fine-tuning other layers or utilizing LoRA. We also find that leveraging more relevant training samples using selective layers does not result in better performance. By combining ELearn and EFit, we create a new model (ELearnFit) that leverages the benefits of both few-shot learning and fine-tuning and produces superior performance to either model alone. We also use ELearnFit to highlight the trade-offs between prompting and fine-tuning, especially for situations where only a limited number of annotated samples are available. Ultimately, our research provides practical techniques to optimize news summarization during the prompting and fine-tuning stages and enhances the synthesis of news articles.

link: http://arxiv.org/abs/2405.02710v1

The Role of AI in Peer Support for Young People: A Study of Preferences for Human- and AI-Generated Responses

Jordyn Young, Laala M Jawara, Diep N Nguyen, Brian Daly, Jina Huh-Yoo, Afsaneh Razi

Generative Artificial Intelligence (AI) is integrated into everyday technology, including news, education, and social media. AI has further pervaded private conversations as conversational partners, auto-completion, and response suggestions. As social media becomes young people's main method of peer support exchange, we need to understand when and how AI can facilitate and assist in such exchanges in a beneficial, safe, and socially appropriate way. We asked 622 young people to complete an online survey and evaluate blinded human- and AI-generated responses to help-seeking messages. We found that participants preferred the AI-generated response to situations about relationships, self-expression, and physical health. However, when addressing a sensitive topic, like suicidal thoughts, young people preferred the human response. We also discuss the role of training in online peer support exchange and its implications for supporting young people's well-being. Disclaimer: This paper includes sensitive topics, including suicide ideation. Reader discretion is advised.

link: http://dx.doi.org/10.1145/3613904.3642574

CoE-SQL: In-Context Learning for Multi-Turn Text-to-SQL with Chain-of-Editions

Hanchong Zhang, Ruisheng Cao, Hongshen Xu, Lu Chen, Kai Yu

Recently, Large Language Models (LLMs) have been demonstrated to possess impressive capabilities in a variety of domains and tasks. We investigate the issue of prompt design in the multi-turn text-to-SQL task and attempt to enhance the LLMs' reasoning capacity when generating SQL queries. In the conversational context, the current SQL query can be modified from the preceding SQL query with only a few operations due to the context dependency. We introduce our method called CoE-SQL which can prompt LLMs to generate the SQL query based on the previously generated SQL query with an edition chain. We also conduct extensive ablation studies to determine the optimal configuration of our approach. Our approach outperforms different in-context learning baselines stably and achieves state-of-the-art performances on two benchmarks SParC and CoSQL using LLMs, which is also competitive to the SOTA fine-tuned models.

link: http://arxiv.org/abs/2405.02712v1

Beyond Relevance: Evaluate and Improve Retrievers on Perspective Awareness

Xinran Zhao, Tong Chen, Sihao Chen, Hongming Zhang, Tongshuang Wu

The task of Information Retrieval (IR) requires a system to identify relevant documents based on users' information needs. In real-world scenarios, retrievers are expected to not only rely on the semantic relevance between the documents and the queries but also recognize the nuanced intents or perspectives behind a user query. For example, when asked to verify a claim, a retrieval system is expected to identify evidence from both supporting vs. contradicting perspectives, for the downstream system to make a fair judgment call. In this work, we study whether retrievers can recognize and respond to different perspectives of the queries -- beyond finding relevant documents for a claim, can retrievers distinguish supporting vs. opposing documents? We reform and extend six existing tasks to create a benchmark for retrieval, where we have diverse perspectives described in free-form text, besides root, neutral queries. We show that current retrievers covered in our experiments have limited awareness of subtly different perspectives in queries and can also be biased toward certain perspectives. Motivated by the observation, we further explore the potential to leverage geometric features of retriever representation space to improve the perspective awareness of retrievers in a zero-shot manner. We demonstrate the efficiency and effectiveness of our projection-based methods on the same set of tasks. Further analysis also shows how perspective awareness improves performance on various downstream tasks, with 4.2% higher accuracy on AmbigQA and 29.9% more correlation with designated viewpoints on essay writing, compared to non-perspective-aware baselines.

link: http://arxiv.org/abs/2405.02714v1

AFter: Attention-based Fusion Router for RGBT Tracking

Andong Lu, Wanyu Wang, Chenglong Li, Jin Tang, Bin Luo

Multi-modal feature fusion as a core investigative component of RGBT tracking emerges numerous fusion studies in recent years. However, existing RGBT tracking methods widely adopt fixed fusion structures to integrate multi-modal feature, which are hard to handle various challenges in dynamic scenarios. To address this problem, this work presents a novel \emph{A}ttention-based \emph{F}usion rou\emph{ter} called AFter, which optimizes the fusion structure to adapt to the dynamic challenging scenarios, for robust RGBT tracking. In particular, we design a fusion structure space based on the hierarchical attention network, each attention-based fusion unit corresponding to a fusion operation and a combination of these attention units corresponding to a fusion structure. Through optimizing the combination of attention-based fusion units, we can dynamically select the fusion structure to adapt to various challenging scenarios. Unlike complex search of different structures in neural architecture search algorithms, we develop a dynamic routing algorithm, which equips each attention-based fusion unit with a router, to predict the combination weights for efficient optimization of the fusion structure. Extensive experiments on five mainstream RGBT tracking datasets demonstrate the superior performance of the proposed AFter against state-of-the-art RGBT trackers. We release the code in https://github.com/Alexadlu/AFter.

link: http://arxiv.org/abs/2405.02717v1

Taming Equilibrium Bias in Risk-Sensitive Multi-Agent Reinforcement Learning

Yingjie Fei, Ruitu Xu

We study risk-sensitive multi-agent reinforcement learning under general-sum Markov games, where agents optimize the entropic risk measure of rewards with possibly diverse risk preferences. We show that using the regret naively adapted from existing literature as a performance metric could induce policies with equilibrium bias that favor the most risk-sensitive agents and overlook the other agents. To address such deficiency of the naive regret, we propose a novel notion of regret, which we call risk-balanced regret, and show through a lower bound that it overcomes the issue of equilibrium bias. Furthermore, we develop a self-play algorithm for learning Nash, correlated, and coarse correlated equilibria in risk-sensitive Markov games. We prove that the proposed algorithm attains near-optimal regret guarantees with respect to the risk-balanced regret.

link: http://arxiv.org/abs/2405.02724v1

A Mathematical Model of the Hidden Feedback Loop Effect in Machine Learning Systems

Andrey Veprikov, Alexander Afanasiev, Anton Khritankov

Widespread deployment of societal-scale machine learning systems necessitates a thorough understanding of the resulting long-term effects these systems have on their environment, including loss of trustworthiness, bias amplification, and violation of AI safety requirements. We introduce a repeated learning process to jointly describe several phenomena attributed to unintended hidden feedback loops, such as error amplification, induced concept drift, echo chambers and others. The process comprises the entire cycle of obtaining the data, training the predictive model, and delivering predictions to end-users within a single mathematical model. A distinctive feature of such repeated learning setting is that the state of the environment becomes causally dependent on the learner itself over time, thus violating the usual assumptions about the data distribution. We present a novel dynamical systems model of the repeated learning process and prove the limiting set of probability distributions for positive and negative feedback loop modes of the system operation. We conduct a series of computational experiments using an exemplary supervised learning problem on two synthetic data sets. The results of the experiments correspond to the theoretical predictions derived from the dynamical model. Our results demonstrate the feasibility of the proposed approach for studying the repeated learning processes in machine learning systems and open a range of opportunities for further research in the area.

link: http://arxiv.org/abs/2405.02726v1

U-DiTs: Downsample Tokens in U-Shaped Diffusion Transformers

Yuchuan Tian, Zhijun Tu, Hanting Chen, Jie Hu, Chao Xu, Yunhe Wang

Diffusion Transformers (DiTs) introduce the transformer architecture to diffusion tasks for latent-space image generation. With an isotropic architecture that chains a series of transformer blocks, DiTs demonstrate competitive performance and good scalability; but meanwhile, the abandonment of U-Net by DiTs and their following improvements is worth rethinking. To this end, we conduct a simple toy experiment by comparing a U-Net architectured DiT with an isotropic one. It turns out that the U-Net architecture only gain a slight advantage amid the U-Net inductive bias, indicating potential redundancies within the U-Net-style DiT. Inspired by the discovery that U-Net backbone features are low-frequency-dominated, we perform token downsampling on the query-key-value tuple for self-attention and bring further improvements despite a considerable amount of reduction in computation. Based on self-attention with downsampled tokens, we propose a series of U-shaped DiTs (U-DiTs) in the paper and conduct extensive experiments to demonstrate the extraordinary performance of U-DiT models. The proposed U-DiT could outperform DiT-XL/2 with only 1/6 of its computation cost. Codes are available at https://github.com/YuchuanTian/U-DiT.

link: http://arxiv.org/abs/2405.02730v1

Systematic Review: Anomaly Detection in Connected and Autonomous Vehicles

J. R. V. Solaas, N. Tuptuk, E. Mariconti

This systematic review focuses on anomaly detection for connected and autonomous vehicles. The initial database search identified 2160 articles, of which 203 were included in this review after rigorous screening and assessment. This study revealed that the most commonly used Artificial Intelligence (AI) algorithms employed in anomaly detection are neural networks like LSTM, CNN, and autoencoders, alongside one-class SVM. Most anomaly-based models were trained using real-world operational vehicle data, although anomalies, such as attacks and faults, were often injected artificially into the datasets. These models were evaluated mostly using five key evaluation metrics: recall, accuracy, precision, F1-score, and false positive rate. The most frequently used selection of evaluation metrics used for anomaly detection models were accuracy, precision, recall, and F1-score. This systematic review presents several recommendations. First, there is a need to incorporate multiple evaluation metrics to provide a comprehensive assessment of the anomaly detection models. Second, only a small proportion of the studies have made their models open source, indicating a need to share models publicly to facilitate collaboration within the research community, and to validate and compare findings effectively. Third, there is a need for benchmarking datasets with predefined anomalies or cyberattacks to test and improve the effectiveness of the proposed anomaly-based detection models. Furthermore, there is a need for future research to investigate the deployment of anomaly detection to a vehicle to assess its performance on the road. There is a notable lack of research done on intrusion detection systems using different protocols to CAN, such as Ethernet and FlexRay.

link: http://arxiv.org/abs/2405.02731v1

Recall Them All: Retrieval-Augmented Language Models for Long Object List Extraction from Long Documents

Sneha Singhania, Simon Razniewski, Gerhard Weikum

Methods for relation extraction from text mostly focus on high precision, at the cost of limited recall. High recall is crucial, though, to populate long lists of object entities that stand in a specific relation with a given subject. Cues for relevant objects can be spread across many passages in long texts. This poses the challenge of extracting long lists from long texts. We present the L3X method which tackles the problem in two stages: (1) recall-oriented generation using a large language model (LLM) with judicious techniques for retrieval augmentation, and (2) precision-oriented scrutinization to validate or prune candidates. Our L3X method outperforms LLM-only generations by a substantial margin.

link: http://arxiv.org/abs/2405.02732v1