Mon 2024.05.13

Heterogeneous Graph Neural Networks with Loss-decrease-aware Curriculum Learning

Yili Wang

In recent years, heterogeneous graph neural networks (HGNNs) have achieved excellent performance in handling heterogeneous information networks (HINs). Curriculum learning is a machine learning strategy where training examples are presented to a model in a structured order, starting with easy examples and gradually increasing difficulty, aiming to improve learning efficiency and generalization. To better exploit the rich information in HINs, previous methods have started to explore the use of curriculum learning strategy to train HGNNs. Specifically, these works utilize the absolute value of the loss at each training epoch to evaluate the learning difficulty of each training sample. However, the relative loss, rather than the absolute value of loss, reveals the learning difficulty. Therefore, we propose a novel loss-decrease-aware training schedule (LDTS). LDTS uses the trend of loss decrease between each training epoch to better evaluating the difficulty of training samples, thereby enhancing the curriculum learning of HGNNs for downstream tasks. Additionally, we propose a sampling strategy to alleviate training imbalance issues. Our method further demonstrate the efficacy of curriculum learning in enhancing HGNNs capabilities. We call our method Loss-decrease-aware Heterogeneous Graph Neural Networks (LDHGNN). The code is public at https://github.com/wangyili00/LDHGNN.

link: http://arxiv.org/abs/2405.06522v1

Prompting Large Language Models with Knowledge Graphs for Question Answering Involving Long-tail Facts

Wenyu Huang, Guancheng Zhou, Mirella Lapata, Pavlos Vougiouklis, Sebastien Montella, Jeff Z. Pan

Although Large Language Models (LLMs) are effective in performing various NLP tasks, they still struggle to handle tasks that require extensive, real-world knowledge, especially when dealing with long-tail facts (facts related to long-tail entities). This limitation highlights the need to supplement LLMs with non-parametric knowledge. To address this issue, we analysed the effects of different types of non-parametric knowledge, including textual passage and knowledge graphs (KGs). Since LLMs have probably seen the majority of factual question-answering datasets already, to facilitate our analysis, we proposed a fully automatic pipeline for creating a benchmark that requires knowledge of long-tail facts for answering the involved questions. Using this pipeline, we introduce the LTGen benchmark. We evaluate state-of-the-art LLMs in different knowledge settings using the proposed benchmark. Our experiments show that LLMs alone struggle with answering these questions, especially when the long-tail level is high or rich knowledge is required. Nonetheless, the performance of the same models improved significantly when they were prompted with non-parametric knowledge. We observed that, in most cases, prompting LLMs with KG triples surpasses passage-based prompting using a state-of-the-art retriever. In addition, while prompting LLMs with both KG triples and documents does not consistently improve knowledge coverage, it can dramatically reduce hallucinations in the generated content.

link: http://arxiv.org/abs/2405.06524v1

Semantic and Spatial Adaptive Pixel-level Classifier for Semantic Segmentation

Xiaowen Ma, Zhenliang Ni, Xinghao Chen

Vanilla pixel-level classifiers for semantic segmentation are based on a certain paradigm, involving the inner product of fixed prototypes obtained from the training set and pixel features in the test image. This approach, however, encounters significant limitations, i.e., feature deviation in the semantic domain and information loss in the spatial domain. The former struggles with large intra-class variance among pixel features from different images, while the latter fails to utilize the structured information of semantic objects effectively. This leads to blurred mask boundaries as well

as a deficiency of fine-grained recognition capability. In this paper, we propose a novel Semantic and Spatial Adaptive (SSA) classifier to address the above challenges. Specifically, we employ the coarse masks obtained from the fixed prototypes as a guide to adjust the fixed prototype towards the center of the semantic and spatial domains in the test image. The adapted prototypes in semantic and spatial domains are then simultaneously considered to accomplish classification decisions. In addition, we propose an online multi-domain distillation learning strategy to improve the adaption process. Experimental results on three publicly available benchmarks show that the proposed SSA significantly improves the segmentation performance of the baseline models with only a minimal increase in computational cost. Code is available at https://github.com/xwmaxwma/SSA.

link: http://arxiv.org/abs/2405.06525v1

Controllable Image Generation With Composed Parallel Token Prediction

Jamie Stirling, Noura Al-Moubayed

Compositional image generation requires models to generalise well in situations where two or more input concepts do not necessarily appear together in training (compositional generalisation). Despite recent progress in compositional image generation via composing continuous sampling processes such as diffusion and energy-based models, composing discrete generative processes has remained an open challenge, with the promise of providing improvements in efficiency, interpretability and simplicity. To this end, we propose a formulation for controllable conditional generation of images via composing the log-probability outputs of discrete generative models of the latent space. Our approach, when applied alongside VQ-VAE and VQ-GAN, achieves state-of-the-art generation accuracy in three distinct settings (FFHQ, Positional CLEVR and Relational CLEVR) while attaining competitive Fr\'echet Inception Distance (FID) scores. Our method attains an average generation accuracy of 80.71% across the studied settings. Our method also outperforms the next-best approach (ranked by accuracy) in terms of FID in seven out of nine experiments, with an average FID of 24.23 (an average improvement of -9.58). Furthermore, our method offers a $2.3\times$ to $12\times$ speedup over comparable continuous compositional methods on our hardware. We find that our method can generalise to combinations of input conditions that lie outside the training data (e.g. more objects per image) in addition to offering an interpretable dimension of controllability via concept weighting. We further demonstrate that our approach can be readily applied to an open pre-trained discrete text-to-image model without any fine-tuning, allowing for fine-grained control of text-to-image generation.

link: http://arxiv.org/abs/2405.06535v1

Mesh Denoising Transformer

Wenbo Zhao, Xianming Liu, Deming Zhai, Junjun Jiang, Xiangyang Ji

Mesh denoising, aimed at removing noise from input meshes while preserving their feature structures, is a practical yet challenging task. Despite the remarkable progress in learning-based mesh denoising methodologies in recent years, their network designs often encounter two principal drawbacks: a dependence on single-modal geometric representations, which fall short in capturing the multifaceted attributes of meshes, and a lack of effective global feature aggregation, hindering their ability to fully understand the mesh's comprehensive structure. To tackle these issues, we propose SurfaceFormer, a pioneering Transformer-based mesh denoising framework. Our first contribution is the development of a new representation known as Local Surface Descriptor, which is crafted by establishing polar systems on each mesh face, followed by sampling points from adjacent surfaces using geodesics. The normals of these points are organized into 2D patches, mimicking images to capture local geometric intricacies, whereas the poles and vertex coordinates are consolidated into a point cloud to embody spatial information. This advancement surmounts the hurdles posed by the irregular and non-Euclidean characteristics of mesh data, facilitating a smooth integration with Transformer architecture. Next, we propose a dual-stream structure consisting of a Geometric Encoder branch and a Spatial Encoder branch, which jointly encode local geometry details and spatial information to fully explore multimodal information for mesh denoising. A subsequent Denoising Transformer module receives the multimodal information and achieves

efficient global feature aggregation through self-attention operators. Our experimental evaluations demonstrate that this novel approach outperforms existing state-of-the-art methods in both objective and subjective assessments, marking a significant leap forward in mesh denoising.

link: http://arxiv.org/abs/2405.06536v1

ATSumm: Auxiliary information enhanced approach for abstractive disaster Tweet Summarization with sparse training data

Piyush Kumar Garg, Roshni Chakraborty, Sourav Kumar Dandapat

The abundance of situational information on Twitter poses a challenge for users to manually discern vital and relevant information during disasters. A concise and human-interpretable overview of this information helps decision-makers in implementing efficient and quick disaster response. Existing abstractive summarization approaches can be categorized as sentence-based or key-phrase-based approaches. This paper focuses on sentence-based approach, which is typically implemented as a dual-phase procedure in literature. The initial phase, known as the extractive phase, involves identifying the most relevant tweets. The subsequent phase, referred to as the abstractive phase, entails generating a more human-interpretable summary. In this study, we adopt the methodology from prior research for the extractive phase. For the abstractive phase of summarization, most existing approaches employ deep learning-based frameworks, which can either be pre-trained or require training from scratch. However, to achieve the appropriate level of performance, it is imperative to have substantial training data for both methods, which is not readily available. This work presents an Abstractive Tweet Summarizer (ATSumm) that effectively addresses the issue of data sparsity by using auxiliary information. We introduced the Auxiliary Pointer Generator Network (AuxPGN) model, which utilizes a unique attention mechanism called Key-phrase attention. This attention mechanism incorporates auxiliary information in the form of key-phrases and their corresponding importance scores from the input tweets. We evaluate the proposed approach by comparing it with 10 state-of-the-art approaches across 13 disaster datasets. The evaluation results indicate that ATSumm achieves superior performance compared to state-of-the-art approaches, with improvement of 4-80% in ROUGE-N F1-score.

link: http://arxiv.org/abs/2405.06541v1

Mitigating Hallucinations in Large Language Models via Self-Refinement-Enhanced Knowledge Retrieval

Mengjia Niu, Hao Li, Jie Shi, Hamed Haddadi, Fan Mo

Large language models (LLMs) have demonstrated remarkable capabilities across various domains, although their susceptibility to hallucination poses significant challenges for their deployment in critical areas such as healthcare. To address this issue, retrieving relevant facts from knowledge graphs (KGs) is considered a promising method. Existing KG-augmented approaches tend to be resource-intensive, requiring multiple rounds of retrieval and verification for each factoid, which impedes their application in real-world scenarios. In this study, we propose Self-Refinement-Enhanced Knowledge Graph Retrieval (Re-KGR) to augment the factuality of LLMs' responses with less retrieval efforts in the medical field. Our approach leverages the attribution of next-token predictive probability distributions across different tokens, and various model layers to primarily identify tokens with a high potential for hallucination, reducing verification rounds by refining knowledge triples associated with these tokens. Moreover, we rectify inaccurate content using retrieved knowledge in the post-processing stage, which improves the truthfulness of generated responses. Experimental results on a medical dataset demonstrate that our approach can enhance the factual capability of LLMs across various foundational models as evidenced by the highest scores on truthfulness.

link: http://arxiv.org/abs/2405.06545v1

Sharp analysis of out-of-distribution error for "importance-weighted" estimators in the overparameterized regime

Kuo-Wei Lai, Vidya Muthukumar

Overparameterized models that achieve zero training error are observed to generalize well on average, but degrade in performance when faced with data that is under-represented in the training sample. In this work, we study an overparameterized Gaussian mixture model imbued with a spurious feature, and sharply analyze the in-distribution and out-of-distribution test error of a cost-sensitive interpolating solution that incorporates "importance weights". Compared to recent work Wang et al. (2021), Behnia et al. (2022), our analysis is sharp with matching upper and lower bounds, and significantly weakens required assumptions on data dimensionality. Our error characterizations also apply to any choice of importance weights and unveil a novel tradeoff between worst-case robustness to distribution shift and average accuracy as a function of the importance weight magnitude.

link: http://arxiv.org/abs/2405.06546v1

OneTo3D: One Image to Re-editable Dynamic 3D Model and Video Generation

Jinwei Lin

One image to editable dynamic 3D model and video generation is novel direction and change in the research area of single image to 3D representation or 3D reconstruction of image. Gaussian Splatting has demonstrated its advantages in implicit 3D reconstruction, compared with the original Neural Radiance Fields. As the rapid development of technologies and principles, people tried to used the Stable Diffusion models to generate targeted models with text instructions. However, using the normal implicit machine learning methods is hard to gain the precise motions and actions control, further more, it is difficult to generate a long content and semantic continuous 3D video. To address this issue, we propose the OneTo3D, a method and theory to used one single image to generate the editable 3D model and generate the targeted semantic continuous time-unlimited 3D video. We used a normal basic Gaussian Splatting model to generate the 3D model from a single image, which requires less volume of video memory and computer calculation ability. Subsequently, we designed an automatic generation and self-adaptive binding mechanism for the object armature. Combined with the re-editable motions and actions analyzing and controlling algorithm we proposed, we can achieve a better performance than the SOTA projects in the area of building the 3D model precise motions and actions control, and generating a stable semantic continuous time-unlimited 3D video with the input text instructions. Here we will analyze the detailed implementation methods and theories analyses. Relative comparisons and conclusions will be presented. The project code is open source.

link: http://arxiv.org/abs/2405.06547v1

Sampling the Swadesh List to Identify Similar Languages with Tree Spaces

Garett Ordway, Vic Patrangenaru

Communication plays a vital role in human interaction. Studying language is a worthwhile task and more recently has become quantitative in nature with developments of fields like quantitative comparative linguistics and lexicostatistics. With respect to the authors own native languages, the ancestry of the English language and the Latin alphabet are of the primary interest. The Indo-European Tree traces many modern languages back to the Proto-Indo-European root. Swadesh's cognates played a large role in developing that historical perspective where some of the primary branches are Germanic, Celtic, Italic, and Balto-Slavic. This paper will use data analysis on open books where the simplest singular space is the 3-spider - a union T3 of three rays with their endpoints glued at a point 0 - which can represent these tree spaces for language clustering. These trees are built using a single linkage method for clustering based on distances between samples from languages which use the Latin Script. Taking three languages at a time, the barycenter is determined. Some initial results have found both non-sticky and sticky sample means. If the mean exhibits non-sticky properties, then one language may come from a different ancestor than the other two. If the mean is considered sticky, then the languages may share a common ancestor or all languages may have different ancestry.

link: http://arxiv.org/abs/2405.06549v1

ADSumm: Annotated Ground-truth Summary Datasets for Disaster Tweet Summarization

Piyush Kumar Garg, Roshni Chakraborty, Sourav Kumar Dandapat

Online social media platforms, such as Twitter, provide valuable information during disaster events. Existing tweet disaster summarization approaches provide a summary of these events to aid government agencies, humanitarian organizations, etc., to ensure effective disaster response. In the literature, there are two types of approaches for disaster summarization, namely, supervised and unsupervised approaches. Although supervised approaches are typically more effective, they necessitate a sizable number of disaster event summaries for testing and training. However, there is a lack of good number of disaster summary datasets for training and evaluation. This motivates us to add more datasets to make supervised learning approaches more efficient. In this paper, we present ADSumm, which adds annotated ground-truth summaries for eight disaster events which consist of both natural and man-made disaster events belonging to seven different countries. Our experimental analysis shows that the newly added datasets improve the performance of the supervised summarization approaches by 8-28% in terms of ROUGE-N F1-score. Moreover, in newly annotated dataset, we have added a category label for each input tweet which helps to ensure good coverage from different categories in summary. Additionally, we have added two other features relevance label and key-phrase, which provide information about the quality of a tweet and explanation about the inclusion of the tweet into summary, respectively. For ground-truth summary creation, we provide the annotation procedure adapted in detail, which has not been described in existing literature. Experimental analysis shows the quality of ground-truth summary is very good with Coverage, Relevance and Diversity.

link: http://arxiv.org/abs/2405.06551v1

Scalable Property Valuation Models via Graph-based Deep Learning

Enrique Riveros, Carla Vairetti, Christian Wegmann, Santiago Truffa, Sebastián Maldonado

This paper aims to enrich the capabilities of existing deep learning-based automated valuation models through an efficient graph representation of peer dependencies, thus capturing intricate spatial relationships. In particular, we develop two novel graph neural network models that effectively identify sequences of neighboring houses with similar features, employing different message passing algorithms. The first strategy consider standard spatial graph convolutions, while the second one utilizes transformer graph convolutions. This approach confers scalability to the modeling process. The experimental evaluation is conducted using a proprietary dataset comprising approximately 200,000 houses located in Santiago, Chile. We show that employing tailored graph neural networks significantly improves the accuracy of house price prediction, especially when utilizing transformer convolutional message passing layers.

link: http://arxiv.org/abs/2405.06553v1

Random matrix theory improved Fréchet mean of symmetric positive definite matrices

Florent Bouchard, Ammar Mian, Malik Tiomoko, Guillaume Ginolhac, Frédéric Pascal

In this study, we consider the realm of covariance matrices in machine learning, particularly focusing on computing Fr\'echet means on the manifold of symmetric positive definite matrices, commonly referred to as Karcher or geometric means. Such means are leveraged in numerous machine-learning tasks. Relying on advanced statistical tools, we introduce a random matrix theory-based method that estimates Fr\'echet means, which is particularly beneficial when dealing with low sample support and a high number of matrices to average. Our experimental evaluation, involving both synthetic and real-world EEG and hyperspectral datasets, shows that we largely outperform state-of-the-art methods.

link: http://arxiv.org/abs/2405.06558v1

Reservoir Computing Benchmarks: a review, a taxonomy, some best practices

Chester Wringe, Martin Trefzer, Susan Stepney

Reservoir Computing is an Unconventional Computation model to perform computation on various different substrates, such as RNNs or physical materials. The method takes a "black-box" approach, training only the outputs of the system it is built on. As such, evaluating the computational capacity of these systems can be challenging. We review and critique the evaluation methods used in the field of Reservoir Computing. We introduce a categorisation of benchmark tasks. We review multiple examples of benchmarks from the literature as applied to reservoir computing, and note their strengths and shortcomings. We suggest ways in which benchmarks and their uses may be improved to the benefit of the reservoir computing community

link: http://arxiv.org/abs/2405.06561v1

What Can Natural Language Processing Do for Peer Review?

Ilia Kuznetsov, Osama Mohammed Afzal, Koen Dercksen, Nils Dycke, Alexander Goldberg, Tom Hope, Dirk Hovy, Jonathan K. Kummerfeld, Anne Lauscher, Kevin Leyton-Brown, Sheng Lu, Mausam, Margot Mieskes, Aurélie Névéol, Danish Pruthi, Lizhen Qu, Roy Schwartz, Noah A. Smith, Thamar Solorio, Jingyan Wang, Xiaodan Zhu, Anna Rogers, Nihar B. Shah, Iryna Gurevych

The number of scientific articles produced every year is growing rapidly. Providing quality control over them is crucial for scientists and, ultimately, for the public good. In modern science, this process is largely delegated to peer review -- a distributed procedure in which each submission is evaluated by several independent experts in the field. Peer review is widely used, yet it is hard, time-consuming, and prone to error. Since the artifacts involved in peer review -- manuscripts, reviews, discussions -- are largely text-based, Natural Language Processing has great potential to improve reviewing. As the emergence of large language models (LLMs) has enabled NLP assistance for many new tasks, the discussion on machine-assisted peer review is picking up the pace. Yet, where exactly is help needed, where can NLP help, and where should it stand aside? The goal of our paper is to provide a foundation for the future efforts in NLP for peer-reviewing assistance. We discuss peer review as a general process, exemplified by reviewing at AI conferences. We detail each step of the process from manuscript submission to camera-ready revision, and discuss the associated challenges and opportunities for NLP assistance, illustrated by existing work. We then turn to the big challenges in NLP for peer review as a whole, including data acquisition and licensing, operationalization and experimentation, and ethical issues. To help consolidate community efforts, we create a companion repository that aggregates key datasets pertaining to peer review. Finally, we issue a detailed call for action for the scientific community, NLP and AI researchers, policymakers, and funding bodies to help bring the research in NLP for peer review forward. We hope that our work will help set the agenda for research in machine-assisted scientific quality control in the age of AI, within the NLP community and beyond.

link: http://arxiv.org/abs/2405.06563v1