Machine Intelligence Research
www.mi-research.net

20(2), April 2023, 145-146
DOI: 10.1007/s11633-023-1431-y

Editorial for Special Issue on Large-scale Pre-training: Data, Models, and Fine-tuning

In recent years, there has been a surge of interest and rapid development in large-scale pre-training due to the explosive growth of both data and model parameters. Large-scale training has achieved impressive performance milestones across a wide range of practical problems, including natural language processing, computer vision, recommendation systems, robotics, and other basic research areas like bioinformatics. Different from early non-neural models and small models that rely heavily on hand-crafted features, statistical methods, and accurate human annotations, neural models can automatically learn low-level distributed representations and high-level latent semantic information from data. However, the huge numbers of parameters in deep neural models may lead to over-fitting and poor generalization, which has prompted massive efforts to exploit how to pre-train large-scale models on large-scale data. As large-scale human annotations are time-consuming and costly, it is impractical to pre-train large-scale models in a fully-supervised manner. In response, the AI community has recently focused on self-supervised learning algorithms and theories, large-scale pre-training paradigms according to data format, large-scale model architecture designs, and downstream applications.

This special issue seeks original and novel contributions toward advancing the theory, architecture, and algorithmic design for large-scale pre-training in machine intelligence as well as downstream applications. The special issue will provide a timely collection of recent advances to benefit the researchers and practitioners working in the broad research fields of machine intelligence, natural language processing, and computer vision. In the end, ten papers are accepted to form this special issue.

The first paper, entitled "Pre-training in Medical Data: A Survey" from Yixuan Qiu et al., provided an in-depth review of pre-training methods applied to medical data. This paper highlighted the significance of pre-training in the medical domain, especially when data is limited. The paper further summarized a large number of related publications and benchmarks, examining how pre-training methods work on medical data. The in-depth review provided a comprehensive overview of the recent advancements in pre-training techniques for medical data and valuable insights for future studies.

The second paper, entitled "Red Alarm for Pre-trained Models: Universal Vulnerability to Neuron-level Backdoor Attacks" from Zhengyan Zhang et al., pointed out that a pre-trained model that has been backdoored can exhibit malicious behaviors in various downstream tasks without prior knowledge of task information. The paper focuses on the neuron-level backdoor attack (NeuBA), a type of attack where the output representations of trigger-embedded samples are restricted to arbitrary predefined values through additional training. This paper further demonstrated the effectiveness of NeuBA in natural language processing and computer vision tasks and explore defense methods to resist NeuBA, with model pruning being a promising technique. This paper highlights the urgent need for better security measures for pre-trained models to protect against such attacks.

The third paper, entitled by "A Study of Using Synthetic Data for Effective Association Knowledge Learning" from Yuchi Liu et al., reviewed the role of large-scale synthetic data in video analysis. Specifically, this paper explored the use of synthetic data for multi-object tracking (MOT) instead of relying on expensive real video data. This work showed that synthetic data can achieve very similar performance on real-world test sets without domain adaptation techniques. This paper further demonstrated the ability 3D engines to simulate motion factors and the insignificant impact of appearance domain gap on the learning of association knowledge and highlighted the strong customization ability of MOTX to assess the impact of motion factors on MOT quantitatively and bring new insights to the community.

The fourth paper, entitled "EVA2.0: Investigating Open-domain Chinese Dialogue Systems with Large-scale Pre-training" from Yuxian Gu et al., discussed key factors in creating a powerful and human-like chatbot in Chinese scenarios. The paper conducted extensive experiments on data quality control, model architecture designs, training approaches, and decoding strategies. The paper further proposed EVA2.0, a large-scale pre-trained open-domain Chinese dialogue model with 2.8 billion parameters, and show through automatic and human evaluations that it outperforms other open-source counterparts. The paper also suggests future research directions for large-scale Chinese open-domain dialogue systems.

The fifth paper, entitled "Multimodal pretraining from Monolingual to Multilingual" from Liang Zhang et al., focused on multilingual multimodal pre-training, and extended existing monolingual models to multilingual models. To tackle this challenge, this paper proposed a pre-training-based model and two generalization-based models and

Editorial
Special Issue on Large-scale Pre-training: Data, Models, and Fine-tuning

incorporated the audio modality. These models achieve state-of-the-art performance on multilingual vision-text retrieval, visual question answering, and image captioning benchmarks.

The sixth paper, entitled "Offline Pre-trained Multi-Agent Decision Transformer" from Linghui Meng et al, explored the potential of multi-agent reinforcement learning (MARL) tasks. Currently, the paradigm of offline pre-training with online fine-tuning has not been studied in MARL, and there are no datasets or benchmarks available for offline MARL research. To close this gap, this paper constructed large-scale datasets and examine the usage of the decision transformer in MARL, and further proposed multi-agent decision transformer with both offline and online MARL tasks.

The seventh paper, entitled "Compositional Prompting Video-language Models to Understand Procedure in Instructional Videos" from Guyue Hu et al., aimed to understand long-term procedures in instructional videos. This paper proposed a compositional prompt learning (CPL) framework that prompts short-term video-language models to understand long-term procedures. The CPL framework consists of one visual prompt and three compositional textual prompts to distill knowledge from short-term models. This paper further highlights the potential for future research in procedure understanding in instructional videos.

The eighth paper, entitled "Mitigating Spurious Correlations for Self-supervised Recommendation" from Xin-Yu Lin et al., discussed the issue of spurious correlations in self-supervised learning (SSL) recommendation models that result in poor generalization. The paper further proposes an invariant feature learning framework to automatically mitigate the effect of spurious correlations, which involves the automatic masking of spurious features without supervision and blocking the negative effect transmission from spurious features during SSL. This work revealed the potential of invariant feature learning for bias mitigation in recommendation systems pre-training.

The ninth paper, entitled "DynamicRetriever: A Pre-trained Model-based IR System Without an Explicit Index" from Yu-Jia Zhou et al., focused on information retrieval (IR) and challenged the index-retrieve-rerank paradigm by replacing the explicit index with a pre-trained model. The proposed IR system, DynamicRetriever, directly returns document identifiers for a given query, enabling the traditional static index to be parameterized with a pre-training model. This paper revealed the potential of the pre-trained model to remove the index construction process for information retrieval.

The last paper, entitled "Vision Enhanced Generative Pre-trained Language Model for Multimodal Sentence Summarization" from Liqiang Jing et al., focused on the problem of multimodal sentence summarization (MMSS) at the intersection between computer vision and natural language processing. This paper pointed out the limitations of existing methods in MMSS that the generative ability of pre-trained language models is not fully exploited in text generation tasks. This paper further proposed Vision-GPLM to address the challenge of injecting visual information into generative language models while preserving their generation ability.

These papers discussed the role of large-scale pre-training from diverse perspectives, including natural language processing, computer vision, multimodal analysis, reinforcement learning, recommendation systems, information retrieval, and medical data. These works provide valuable insights, methodologies, and techniques that contribute to the advancement of large-scale pre-training. We hope our readers will enjoy reading and learning from these cutting-edge works.

Ji-Rong Wen
Renmin University of China, *China*
Zi Huang
The University of Queensland, *Australia*
Hanwang Zhang
Nanyang Technological University, *Singapore*

Ji-Rong Wen received the Ph. D. degree in computer science from Chinese Academy of Science, China in 1999. He once was a senior researcher and group manager at Microsoft Research Asia (MSRA), China. He is a full professor, the dean of School of Information, and the executive dean of Gaoling School of Artificial Intelligence, Renmin University of China, China. He is the PC Chair of SIGIR 2020 and the Associate Editor of ACM TOIS and IEEE TKDE.

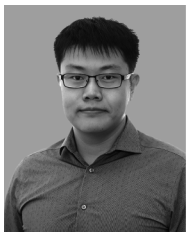
His main research interests include information retrieval, data mining, and machine learning.

E-mail: jrwen@ruc.edu.cn (Corresponding author)
ORCID iD: 0000-0002-9777-9676

Zi Huang received the B. Sc. degree in computer science from Tsinghua University, China in 2001, and the Ph. D. degree in computer science from School of Information Technology and Electrical Engineering, The University of Queensland, Australia in 2007. She is a professor and ARC Future Fellow in School of Information Technology and Electrical Engineering, The University of Queensland, Australia.

Her research interests include multimedia indexing and search, social data analysis, and knowledge discovery.

E-mail: huang@itee.uq.edu.au
ORCID iD: 0000-0002-9738-4949



Hanwang Zhang received the B. Eng. (Hons.) degree in computer science from Zhejiang University, China in 2009, and the Ph. D. degree in computer science from the National University of Singapore, Singapore in 2014. He is currently an associate professor at Nanyang Technological University, Singapore.

His research interests include computer vision, multimedia, and social media.

E-mail: hanwangzhang@ntu.edu.sg
ORCID iD: 0000-0001-7374-8739