



Workshop on Model Mining

Shan You
SenseTime Research
Beijing, China
youshan@sensetime.com

Fei Wang
SenseTime Research
Beijing, China
wangfei@senseauto.com

Chang Xu
School of Computer Science, Faculty of Engineering
The University of Sydney
Sydney, Australia
c.xu@sydney.edu.au

Changshui Zhang
Institute for Artificial Intelligence, Tsinghua University
(THUAI)
Beijing National Research Center for Information Science
and Technology (BNRist)
Department of Automation, Tsinghua University
Beijing, China
zcs@mail.tsinghua.edu.cn

ABSTRACT

How to **mine the knowledge in the pretrained models** is of significance in achieving more promising performance, since practitioners have access to many pretrained models easily. This **Workshop on Model Mining** aims to investigate more diverse and advanced manners in mining knowledge within models, which tends to leverage the pretrained models more wisely, elegantly and systematically. There are many topics related to this workshop, such as distilling a lightweight model from a well-trained heavy model via teacher-student paradigm, and boosting the performance of the model by carefully designing the predecessor tasks, e.g., pre-training, self-supervised and contrastive learning. Model mining as a special way of data mining is relevant to SIGKDD, and its audience including researchers and engineers will benefit a lot for designing more advanced algorithms for their tasks.

CCS CONCEPTS

• **Computing methodologies** → **Knowledge representation and reasoning**; **Computer vision representations**; **Machine learning algorithms**; **Neural networks**.

KEYWORDS

knowledge distillation, large-scale pretraining, model mining, transfer learning

ACM Reference Format:

Shan You, Chang Xu, Fei Wang, and Changshui Zhang. 2021. Workshop on Model Mining. In *Proceedings of the 27th ACM SIGKDD Conference on Knowledge Discovery and Data Mining (KDD '21), August 14–18, 2021, Virtual Event, Singapore*. ACM, New York, NY, USA, 2 pages. <https://doi.org/10.1145/3447548.3469471>

Permission to make digital or hard copies of part or all of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for third-party components of this work must be honored. For all other uses, contact the owner/author(s).

KDD '21, August 14–18, 2021, Virtual Event, Singapore.

© 2021 Copyright held by the owner/author(s).

ACM ISBN 978-1-4503-8332-5/21/08.

<https://doi.org/10.1145/3447548.3469471>

1 INTRODUCTION AND MOTIVATION

Recent years have witnessed the prosperity of deep learning in a variety of applications, such as computer vision, natural language processing and speech recognition and enhancement. And this success stimulates the proliferation of many benchmark deep neural models. Before practitioners develop models for the task of interest, they can actually have access to some off-the-shelf models, which are already trained on the same, similar or even totally different tasks. Those pretrained models are usually dedicated to mine the training data well, and the informative knowledge is supposed to exist in the learned representations or the model weights themselves. In this sense, how to mine the knowledge in the pretrained models is also of significance in achieving more promising performance.

There are multiple ways to fulfill the model mining. For example, inspired by the human education, we can leverage the pretrained models as privileged information [8] or additional resources following the teacher-student paradigm [4–6]. Then the knowledge can be manifested and distilled on the learned logits or intermediate features, such as attention maps [10], similarity among examples [9], and local graphs [7]. As a result, the student models are expected to perform better with the help of the mined knowledge in teacher models. Likewise, pretrained models can also serve as a general yet powerful platform for adapting customized models for different target tasks. For example, large-scale pretrained models with huge parameters as BERT [2], GPT-3 [1], and ViT [3] show great generalization ability in many downstream tasks, and we can have decent performance for the target task with only limited data.

2 OBJECTIVE AND TOPICS OF INTEREST

The TOPIC model mining of this workshop aims to investigate more diverse and advanced manners in mining knowledge within models, which tends to leverage the pretrained models more wisely, elegantly and systematically. The topics of interest include the theory, algorithm and application of mining the values of models in different scenarios

- Distilling a lightweight model from a well-trained heavy model, e.g., knowledge distillation or the teacher-student paradigm;

- Continuously upgrading the model with new data, new concept or new tasks, e.g., continual learning, lifelong learning, zero-shot learning and multi-task learning.
- Boosting the performance of the model by carefully designing the predecessor tasks, e.g., pre-training, self-supervised and contrastive learning.
- Transferring the models across different datasets, domains or modalities, e.g., transfer learning and domain adaptation.

Model mining as a special way of data mining is relevant to SIGKDD, and its audience including researchers and engineers will benefit a lot for designing more advanced algorithms for their tasks.

3 PROGRAM SKETCH

We plan a half-day workshop. The workshop will consist of invited talks, a panel discussion (by invited speakers), and peer-reviewed papers (oral and poster). Based on the workshop discussion, a follow-up will be to produce a positioning paper that will sketch the base for future promising research on model mining.

4 ORGANIZATION COMMITTEE

Shan You (<https://shanyou92.github.io>) is currently a Senior Researcher at SenseTime, and also a post doc at Tsinghua University. Before that, he received a Bachelor of mathematics and applied mathematics (elite class) from Xi'an Jiaotong University, and a Ph.D. degree of computer science from Peking University. During pursuing his PhD degree, he visited the UBTech Sydney AI Institute, the University of Sydney, and received awards from Qualcomm. His research interests include AutoML, reinforcement learning and other computer vision and machine learning topics, such as face recognition and object tracking. He has published his research outcomes in many top tier conferences and transactions.

Chang Xu (<http://changxu.xyz>) is a Senior Lecturer and ARC DECRA Fellow at the School of Computer Science, University of Sydney. His recent interests focus on efficient deep learning and sparse networks, robust learning under noisy data or labels, and neural architecture search. He has published over 100 papers in prestigious journals and top tier conferences. He has received several paper awards, including Distinguished Paper Award in IJCAI 2018. He regularly served as the senior PC or PC for many conferences, e.g., ICLR, SIGKDD, NIPS, ICML, CVPR, ICCV, IJCAI and AAAI. He has been recognized as Top Ten Distinguished Senior PC Member in IJCAI 2017.

Fei Wang (<http://wangfei.info>) is the Research Director of SenseTime. He leads a vibrant team to do fundamental research of computer vision and develop comprehensive solutions for the mobile intelligence industry. The goal of his team is to establish AI systems that make edge devices more intelligent and efficient. His research interests include Deep Learning, Edge Computation, etc. He has gained over 1000 Google Scholar Citations with recent publications during the last few years. Fei obtained his Bachelor's degree and Master's degree from Beijing University of Posts and Telecommunications.

Changshui Zhang (<http://bigeye.au.tsinghua.edu.cn/english/Introduction.html>) received the B.E. degree in mathematics from Peking University, Beijing, China, in 1986, and the M.S. and Ph.D.

degrees in control science and engineering from Tsinghua University, Beijing, in 1989 and 1992, respectively. In 1992, he joined the Department of Automation, Tsinghua University, where he is currently a professor. His research interests include pattern recognition and machine learning. He is a Fellow member of the IEEE.

5 PROGRAM COMMITTEE

We appreciate the reviewers' efforts and would like to thank the members of the program committee for their valuable support:

- Dr. Erkun Yang, Post doc, University of North Carolina at Chapel Hill, USA
- Dr. Jianlong Wu, Assistant Professor, Shandong University, China
- Dr. Chen Qian, Executive Research Director, SenseTime, China
- Dr. Chaoyue Wang, Postdoc Fellow, University of Sydney, Australia
- Dr. Yuxuan Du, University of Sydney, Australia
- Dr. Hongyang Zhang, Postdoc Fellow, TTIC, USA
- Mr. Fanyun Wei, Researcher, MSRA

And we sincerely thank Mingkai Zheng for the technical support for the official website.

ACKNOWLEDGMENTS

We would like to thank the organizers of the SIGKDD 2021 conference for agreeing to host our workshop and for their support, and also the anonymous reviewers who evaluated our workshop proposal, for their support and constructive evaluation.

REFERENCES

- [1] Tom B Brown, Benjamin Mann, Nick Ryder, Melanie Subbiah, Jared Kaplan, Prafulla Dhariwal, Arvind Neelakantan, Pranav Shyam, Girish Sastry, Amanda Askell, et al. 2020. Language models are few-shot learners. *arXiv preprint arXiv:2005.14165* (2020).
- [2] Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. 2018. Bert: Pre-training of deep bidirectional transformers for language understanding. *arXiv preprint arXiv:1810.04805* (2018).
- [3] Alexey Dosovitskiy, Lucas Beyer, Alexander Kolesnikov, Dirk Weissenborn, Xi-aohua Zhai, Thomas Unterthiner, Mostafa Dehghani, Matthias Minderer, Georg Heigold, Sylvain Gelly, et al. 2020. An image is worth 16x16 words: Transformers for image recognition at scale. *arXiv preprint arXiv:2010.11929* (2020).
- [4] Shangchen Du, Shan You, Xiaojie Li, Jianlong Wu, Fei Wang, Chen Qian, and Changshui Zhang. 2020. Agree to Disagree: Adaptive Ensemble Knowledge Distillation in Gradient Space. *Advances in Neural Information Processing Systems* 33 (2020).
- [5] Geoffrey Hinton, Oriol Vinyals, and Jeff Dean. 2015. Distilling the knowledge in a neural network. *arXiv preprint arXiv:1503.02531* (2015).
- [6] Shumin Kong, Tianyu Guo, Shan You, and Chang Xu. 2020. Learning student networks with few data. In *Proceedings of the AAAI Conference on Artificial Intelligence*, Vol. 34. 4469–4476.
- [7] Xiaojie Li, Jianlong Wu, Hongyu Fang, Yue Liao, Fei Wang, and Chen Qian. 2020. Local correlation consistency for knowledge distillation. In *European Conference on Computer Vision*. Springer, 18–33.
- [8] David Lopez-Paz, Léon Bottou, Bernhard Schölkopf, and Vladimir Vapnik. 2015. Unifying distillation and privileged information. *arXiv preprint arXiv:1511.03643* (2015).
- [9] Shan You, Chang Xu, Chao Xu, and Dacheng Tao. 2017. Learning from multiple teacher networks. In *Proceedings of the 23rd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. 1285–1294.
- [10] Sergey Zagoruyko and Nikos Komodakis. 2016. Paying more attention to attention: Improving the performance of convolutional neural networks via attention transfer. *arXiv preprint arXiv:1612.03928* (2016).