SustaiNLP 2022

The Third Workshop on Simple and Efficient Natural Language Processing

Proceedings of the Workshop

December 7, 2022

Order copies of this and other ACL proceedings from:

Association for Computational Linguistics (ACL)
209 N. Eighth Street
Stroudsburg, PA 18360
USA
Tel: +1-570-476-8006
Fax: +1-570-476-0860
acl@aclweb.org

ISBN 978-1-959429-24-1

Introduction

It is our great pleasure to welcome you to the third edition of SustaiNLP: Workshop on Simple and Efficient Natural Language Processing.

The Natural Language Processing community has, in recent years, demonstrated a notable focus on improving higher scores on standard benchmarks and taking the lead on community-wide leaderboards (e.g., GLUE, SentEval). While this aspiration has led to improvements in benchmark performance of (predominantly neural) models, it has also came at a cost, i.e., increased model complexity and the ever-growing amount of computational resources required for training and using the current state-of-the-art models. Moreover, the recent research efforts have, for the most part, failed to identify sources of empirical gains in models, often failing to empirically justify the model complexity beyond benchmark performance.

Because of these easily observable trends, we organized the SustaiNLP workshop with the goal of promoting more sustainable NLP research and practices, with two main objectives: (1) encouraging development of more efficient NLP models; and (2) providing simpler architectures and empirical justification of model complexity. For both aspects, we encouraged submissions from all topical areas of NLP.

Besides the original research papers (short and long), we encouraged cross-submissions of work that has been published at other events as well as extended abstracts of work in progress that fit the scope and aims of the workshop (only the original research papers, however, are included in these workshop proceedings).

This year, we received 17 submissions from ARR, proposing a multitude of viable resource-efficient NLP methods and spanning a wide range of NLP applications. We have selected 8 submissions for presentation at the workshop, yielding an acceptance rate of 47%.

Many thanks to the ARR program committee and our senior area chairs for their thorough and thoughtful reviews. We would also like to thank to our panelists and invited speakers whose discussions and talks we strongly believe will make the workshop exciting and memorable.

We are looking forward to the third edition of the SustaiNLP workshop!

SustaiNLP Organizers
November 2022

Organizing Committee

Organizers

Angela Fan, INRIA Nancy and Facebook AI Research
Iryna Gurevych, TU Darmstadt
Yufang Hou, IBM Research Ireland
Zornitsa Kozareva, Facebook AI Research
Sasha Luccioni, HuggingFace Inc.
Nafise Sadat Moosavi, University of Sheffield
Sujith Ravi, SliceX AI
Gyuwan Kim, UC Santa Barbara
Roy Schwartz, Hebrew University of Jerusalem
Andreas Rücklé, Amazon Search Berlin

Program Committee

Senior Area Chairs

Daniil Sorokin, Amazon Development Center Germany
Leon Derczynski, IT University
Diego Marcheggiani, Amazon
Nishant Subramani, Allen Institute for Artificial Intelligence
Gabriel Stanovsky, Hebrew University of Jerusalem
Emma Strubell, Carnegie Mellon University and Google

Invited Speakers

Kurt Keutzer, UC Berkeley
Percy Liang, Stanford University
Hinrich Schütze, University of Munich
Song Han, MIT EECS

Panelists

Kurt Keutzer, UC Berkeley
Percy Liang, Stanford University
Hinrich Schütze, University of Munich
Sam Bowman, New York University
Barbara Plank, University of Munich
Scott Wen-tau Yih, Meta AI - FAIR

Table of Contents

Program

Wednesday, December 7, 2022

09:00 - 10:30 *Opening Remarks and Gather Town Session 1*

Who Says Elephants Can't Run: Bringing Large Scale MoE Models into Cloud Scale Production
Young Jin Kim, Rawn Henry, Raffy Fahim and Hany Hassan

AfroLM: A Self-Active Learning-based Multilingual Pretrained Language Model for 23 African Languages
Bonaventure F. P. Dossou, Atnafu Lambebo Tonja, Oreen Yousuf, Salomey Osei, Abigail Oppong, Iyanuoluwa Shode, Oluwabusayo Olufunke Awoyomi and Chris Chinenye Emezue

Data-Efficient Auto-Regressive Document Retrieval for Fact Verification
James Thorne

AlphaTuning: Quantization-Aware Parameter-Efficient Adaptation of Large-Scale Pre-Trained Language Models
Se Jung Kwon, Jeonghoon Kim, Jeongin Bae, Kang Min Yoo, Jin-Hwa Kim, Baeseong Park, Byeongwook Kim, Jung-Woo Ha, Nako Sung and Dongsoo Lee

Towards Fair Dataset Distillation for Text Classification
Xudong Han, Aili Shen, Yitong Li, Lea Frermann, Timothy Baldwin and Trevor Cohn

Mask More and Mask Later: Efficient Pretraining of Masked Language Models by Disentangling the [MASK] Token
Baohao Liao, David Thulke, Sanjika Hewavitharana, Hermann Ney and Christof Monz

Look Ma, Only 400 Samples! Revisiting the Effectiveness of Automatic N-Gram Rule Generation for Spelling Normalization in Filipino
Lorenzo Jaime Yu Flores and Dragomir Radev

AutoCAD: Automatically Generate Counterfactuals for Mitigating Shortcut Learning
Jiaxin Wen, Yeshuang Zhu, Jinchao Zhang, Jie Zhou and Minlie Huang

Contrastive Demonstration Tuning for Pre-trained Language Models
Xiaozhuan Liang, Ningyu Zhang, Siyuan Cheng, Zhenru Zhang, Chuanqi Tan and Huajun Chen

Reconciliation of Pre-trained Models and Prototypical Neural Networks in Few-shot Named Entity Recognition
Youcheng Huang, Wenqiang Lei, Jie Fu and Jiancheng Lv

Improving the Sample Efficiency of Prompt Tuning with Domain Adaptation
Xu Guo, Boyang Li and Han Yu

Partitioned Gradient Matching-based Data Subset Selection for Compute-Efficient Robust ASR Training
Ashish Mittal, Durga Sivasubramanian, Rishabh Iyer, Preethi Jyothi and Ganesh Ramakrishnan

Thinking about GPT-3 In-Context Learning for Biomedical IE? Think Again
Bernal Jimenez Gutierrez, Nikolas McNeal, Clayton Washington, You Chen, Lang Li, Huan Sun and Yu Su

Summarization as Indirect Supervision for Relation Extraction
Keming Lu, I-Hung Hsu, Wenxuan Zhou, Mingyu Derek Ma and Muhao Chen

Bridging the Training-Inference Gap for Dense Phrase Retrieval
Gyuwan Kim, Jinhyuk Lee, Barlas Oguz, Wenhan Xiong, Yizhe Zhang, Yashar Mehdad and William Yang Wang

Ensemble Transformer for Efficient and Accurate Ranking Tasks: an Application to Question Answering Systems
Yoshitomo Matsubara, Luca Soldaini, Eric Lind and Alessandro Moschitti

Plug-and-Play VQA: Zero-shot VQA by Conjoining Large Pretrained Models with Zero Training
Anthony Meng Huat Tiong, Junnan Li, Boyang Li, Silvio Savarese and Steven C.H. Hoi

Train Flat, Then Compress: Sharpness-Aware Minimization Learns More Compressible Models
Clara Na, Sanket Vaibhav Mehta and Emma Strubell

Sparse Mixers: Combining MoE and Mixing to build a more efficient BERT
James Lee-Thorp and Joshua Ainslie

XDoc: Unified Pre-training for Cross-Format Document Understanding
Jingye Chen, Tengchao Lv, Lei Cui, Cha Zhang and Furu Wei

Scaling Laws Under the Microscope: Predicting Transformer Performance from Small Scale Experiments
Maor Ivgi, Yair Carmon and Jonathan Berant

Wednesday, December 7, 2022 (continued)

11:00 - 12:00 *Oral Presentation 1*

Quadapter: Adapter for GPT-2 Quantization
Minseop Park, Jaeseong You, Markus Nagel and Simyung Chang

AfroLM: A Self-Active Learning-based Multilingual Pretrained Language Model for 23 African Languages
Bonaventure F. P. Dossou, Atnafu Lambebo Tonja, Oreen Yousuf, Salomey Osei, Abigail Oppong, Iyanuoluwa Shode, Oluwabusayo Olufunke Awoyomi and Chris Chinenye Emezue

Who Says Elephants Can't Run: Bringing Large Scale MoE Models into Cloud Scale Production
Young Jin Kim, Rawn Henry, Raffy Fahim and Hany Hassan

Effective Pretraining Objectives for Transformer-based Autoencoders
Luca Di Liello, Matteo Gabburo and Alessandro Moschitti

14:30 - 15:00 *Invited Talk (Hinrich Schutze)*

15:00 - 15:30 *Oral Presentation 2*

Algorithmic Diversity and Tiny Models: Comparing Binary Networks and the Fruit Fly Algorithm on Document Representation Tasks
Tanise Ceron, Nhut Truong and Aurelie Herbelot

Scaling Laws Under the Microscope: Predicting Transformer Performance from Small Scale Experiments
Maor Ivgi, Yair Carmon and Jonathan Berant

16:00 - 17:30 *Gather Town Session 2*

Efficient Two-Stage Progressive Quantization of BERT
Charles Le, Arash Ardakani, Amir Ardakani, Hang Zhang, Yuyan Chen, James J. Clark, Brett H. Meyer and Warren J. Gross

KGRefiner: Knowledge Graph Refinement for Improving Accuracy of Translational Link Prediction Methods
Mohammad Javad Saeedizade, Najmeh Torabian and Behrouz Minaei-Bidgoli

Wednesday, December 7, 2022 (continued)

Generating Multiple-Length Summaries via Reinforcement Learning for Unsupervised Sentence Summarization
Dongmin Hyun, Xiting Wang, Chayoung Park, Xing Xie and Hwanjo Yu

Explore Unsupervised Structures in Pretrained Models for Relation Extraction
Xi Yang, Tao Ji and Yuanbin Wu

Improving Generalization of Pre-trained Language Models via Stochastic Weight Averaging
Phillippe Langlais, Ali Ghodsi, Ahmad Rashid, Mehdi Rezagholizadeh, Ivan Kobyzev and Peng Lu

Continuation KD: Improved Knowledge Distillation through the Lens of Continuation Optimization
Ali Ghodsi, Pascal Poupart, Mehdi Rezagholizadeh, Ivan Kobyzev and Aref Jafari

Effective Pretraining Objectives for Transformer-based Autoencoders
Luca Di Liello, Matteo Gabburo and Alessandro Moschitti

19:00 - 20:00 *Invited Talk (Song Han)*

20:00 - 20:30 *Panel Discussion*

21:00 - 21:30 *Invited Talk (Percy Liang)*

21:30 - 22:00 *Invited Talk (Kurt Keutzer)*

22:00 - 22:30 *Best Paper Awards and Closing Remarks*