SustaiNLP 2020

SustaiNLP: Workshop on Simple and Efficient Natural Language Processing

Proceedings of the Workshop

November 20, 2020
Online workshop

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 978-1-952148-77-4

Introduction

It is our great pleasure to welcome you to the first 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 scores on standard benchmarks and taking the lead on community-wide leaderboards such as (Super)GLUE, SentEval or XTREME. While this led to improvements in benchmark performance of (predominantly neural) models, it also resulted in a worrysome increase in model complexity and the amount of computational resources required for training and using the current state-of-the-art models. Moreover, recent research efforts often fail to identify sources of performance gains in models and to justify model complexity beyond benchmark performance.

Because of these trends as well as the worrysome carbon footprint of (pre)training large neural models, we organized SustaiNLP in order to promote simpler and more sustainable NLP research and practices, with two main objectives: (1) encouraging development of more resource-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 first edition of the workshop also included a *shared task* encouraging an optimal trade-off between the model performance and efficiency during inference. The shared task focused on inference efficiency as it can be difficult to fairly evaluate training efficiency in the most general setting. Moreover, as large-scale pretrained models reach production, it is the computational cost of inference that will account for most of the cumulative lifetime environmental cost of these models.

We received overwhelming 48 submissions (and 1 shared task system description), proposing a multitude of viable resource-efficient NLP methods and spanning a wide range of NLP applications. We have selected 28 submissions for presentation at the workshop, yielding an acceptance rate of 58%). Additionally, the workshop will include presentations of 38 papers accepted for publication in EMNLP Findings, the content of which we judged to be in line with the scope and aims of the workshop.

Many thanks to our program committee 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 first edition of the SustaiNLP workshop!

SustaiNLP Organizers October 2020

Organizers:

Nafise Sadat Moosavi, TU Darmstadt
Angela Fan, INRIA Nancy & Facebook AI Research Paris
Goran Glavaš, University of Mannheim
Vered Shwartz, Allen Institute for AI (AI2)
Shafiq Joty, Nanyang Technological University
Alex Wang, New York University
Thomas Wolf, Huggingface Inc.

Steering Committee:

Sam Bowman, New York University
Mona Diab, George Washington University
Andrew McCallum, University of Massachusetts Amherst
Alexander Rush, Cornell University
Luke Zettlemoyer, University of Washington & Facebook

Program Committee:

Eneko Agirre, Mikel Artetxe, Dennis Aumiller, Guy Boudoukh, Samuel Cahyawijaya, Elizabeth Clark, Alexis Conneau, Rotem Dror, Gerard Dupont, Nouha Dziri, Kiril Gashteovski, Sebastian Gehrmann, Marjan Ghazvininejad, Samujjwal Ghosh, Andreas Hanselowski, Benjamin Heinzerling, Ari Holtzman, Ozan Irsoy, Srinivasan Iyer, Peter Izsak, Mandar Joshi, Ehsan Kamalloo, Gyuwan Kim, Young Jin Kim, Guillaume Lample, Phong Le, Ji-Ung Lee, Louis Martin, Seyed Abolghasem Mirroshandel, Marius Mosbach, Myle Ott, Daraksha Parveen, Matthew Peters, Jonas Pfeiffer, Mohammad Taher Pilehvar, Barbara Plank, Simone Paolo Ponzetto, Ofir Press, Hannah Rashkin, Siva Reddy, Ines Rehbein, Andreas Rücklé, Victor Sanh, Roy Schwartz, Edwin Simpson, Gabriel Stanovsky, Anders Søgaard, Urmish Thakker, Prasetya Ajie Utama, Ivan Vulić, Moshe Wasserblat, Genta Indra Winata, Sam Wiseman, Caiming Xiong, Canwen Xu

Invited Speakers & Panelists:

Mona Diab, George Washington University
Heng Ji, University of Illinois at Urbana-Champaign
Graham Neubig, Carnegie Mellon University
Alexander Rush, Cornell University
Emma Strubell, Carnegie Mellon University
Armand Joulin, Facebook AI Research
Kyunghyun Cho, New York University
Yejin Choi, University of Washington & Allen Institute for Artificial Intelligence (AI2)
Yoav Goldberg, Bar Ilan University
Iryna Gurevych, TU Darmstadt

Table of Contents

Workshop Program

Knowing Right from Wrong: Should We Use More Complex Models for Automatic Short-Answer Scoring in Bahasa Indonesia?
Ali Akbar Septiandri, Yosef Ardhito Winatmoko and Ilham Firdausi Putra

Rank and run-time aware compression of NLP Applications
Urmish Thakker, Jesse Beu, Dibakar Gope, Ganesh Dasika and Matthew Mattina

Incremental Neural Coreference Resolution in Constant Memory
Patrick Xia, João Sedoc and Benjamin Van Durme

Learning Informative Representations of Biomedical Relations with Latent Variable Models
Harshil Shah and Julien Fauqueur

End to End Binarized Neural Networks for Text Classification
Kumar Shridhar, Harshil Jain, Akshat Agarwal and Denis Kleyko

Large Product Key Memory for Pre-trained Language Models
Gyuwan Kim and Tae Hwan Jung

P-SIF: Document Embeddings using Partition Averaging
Vivek Gupta, Ankit Saw, Pegah Nokhiz, Praneeth Netrapalli, Piyush Rai and Partha Talukdar

Exploring the Boundaries of Low-Resource BERT Distillation
Moshe Wasserblat, Oren Pereg and Peter Izsak

Efficient Estimation of Influence of a Training Instance
Sosuke Kobayashi, Sho Yokoi, Jun Suzuki and Kentaro Inui

Efficient Inference For Neural Machine Translation
Yi-Te Hsu, Sarthak Garg, Yi-Hsiu Liao and Ilya Chatsviorkin

Sparse Optimization for Unsupervised Extractive Summarization of Long Documents with the Frank-Wolfe Algorithm
Alicia Tsai and Laurent El Ghaoui

Don't Read Too Much Into It: Adaptive Computation for Open-Domain Question Answering
Yuxiang Wu, Pasquale Minervini, Pontus Stenetorp and Sebastian Riedel

November 20, 2020 (continued)

November 20, 2020 (continued)

A comparison between CNNs and WFAs for Sequence Classification
Ariadna Quattoni and Xavier Carreras

Counterfactual Augmentation for Training Next Response Selection
Seungtaek Choi, Myeongho Jeong, Jinyoung Yeo and Seung-won Hwang

Do We Need to Create Big Datasets to Learn a Task?
Swaroop Mishra and Bhavdeep Singh Sachdeva

Guiding Attention for Self-Supervised Learning with Transformers
Ameet Deshpande and Karthik Narasimhan

Besides the papers listed above, 38 **EMNLP Findings** papers will be presented at SustaiNLP as well.