Workshop on Simple and Efficient Natural Language Processing (SustaiNLP 2020)

Online
20 November 2020

ISBN: 978-1-7138-2005-5

Printed from e-media with permission by:

Curran Associates, Inc.
57 Morehouse Lane
Red Hook, NY 12571

Some format issues inherent in the e-media version may also appear in this print version.

Copyright© (2020) by the Association for Computational Linguistics
All rights reserved.

Printed with permission by Curran Associates, Inc. (2021)

For permission requests, please contact the Association for Computational Linguistics
at the address below.

Association for Computational Linguistics
209 N. Eighth Street
Stroudsburg, Pennsylvania 18360

Phone: 1-570-476-8006
Fax: 1-570-476-0860

acl@aclweb.org

Additional copies of this publication are available from:

Curran Associates, Inc.
57 Morehouse Lane
Red Hook, NY 12571 USA
Phone: 845-758-0400
Fax: 845-758-2633
Email: curran@proceedings.com
Web: www.proceedings.com

Table of Contents

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 1

Rank and run-time aware compression of NLP Applications
Urmish Thakker, Jesse Beu, Dibakar Gope, Ganesh Dasika and Matthew Mattina 8

Learning Informative Representations of Biomedical Relations with Latent Variable Models
Harshil Shah and Julien Fauqueur 19

End to End Binarized Neural Networks for Text Classification
Kumar Shridhar, Harshil Jain, Akshat Agarwal and Denis Kleyko 29

Exploring the Boundaries of Low-Resource BERT Distillation
Moshe Wasserblat, Oren Pereg and Peter Izsak 35

Efficient Estimation of Influence of a Training Instance
Sosuke Kobayashi, Sho Yokoi, Jun Suzuki and Kentaro Inui 41

Efficient Inference For Neural Machine Translation
Yi-Te Hsu, Sarthak Garg, Yi-Hsiu Liao and Ilya Chatsviorkin 48

Sparse Optimization for Unsupervised Extractive Summarization of Long Documents with the Frank-Wolfe Algorithm
Alicia Tsai and Laurent El Ghaoui 54

Don't Read Too Much Into It: Adaptive Computation for Open-Domain Question Answering
Yuxiang Wu, Pasquale Minervini, Pontus Stenetorp and Sebastian Riedel 63

A Two-stage Model for Slot Filling in Low-resource Settings: Domain-agnostic Non-slot Reduction and Pretrained Contextual Embeddings
Cennet Oguz and Ngoc Thang Vu 73

Early Exiting BERT for Efficient Document Ranking
Ji Xin, Rodrigo Nogueira, Yaoliang Yu and Jimmy Lin 83

Keyphrase Generation with GANs in Low-Resources Scenarios
Giuseppe Lancioni, Saida S.Mohamed, Beatrice Portelli, Giuseppe Serra and Carlo Tasso 89

Quasi-Multitask Learning: an Efficient Surrogate for Obtaining Model Ensembles
Norbert Kis-Szabó and Gábor Berend 97

A Little Bit Is Worse Than None: Ranking with Limited Training Data
Xinyu Zhang, Andrew Yates and Jimmy Lin 107

Predictive Model Selection for Transfer Learning in Sequence Labeling Tasks
Parul Awasthy, Bishwaranjan Bhattacharjee, John Kender and Radu Florian 113

Load What You Need: Smaller Versions of Mutlilingual BERT
Amine Abdaoui, Camille Pradel and Grégoire Sigel 119

SqueezeBERT: What can computer vision teach NLP about efficient neural networks?
Forrest Iandola, Albert Shaw, Ravi Krishna and Kurt Keutzer 124

Analysis of Resource-efficient Predictive Models for Natural Language Processing
Raj Pranesh and Ambesh Shekhar 136

Towards Accurate and Reliable Energy Measurement of NLP Models
Qingqing Cao, Aruna Balasubramanian and Niranjan Balasubramanian 141

FastFormers: Highly Efficient Transformer Models for Natural Language Understanding
Young Jin Kim and Hany Hassan 149

A comparison between CNNs and WFAs for Sequence Classification
Ariadna Quattoni and Xavier Carreras 159

Counterfactual Augmentation for Training Next Response Selection
Seungtaek Choi, Myeongho Jeong, Jinyoung Yeo and Seung-won Hwang 164

Do We Need to Create Big Datasets to Learn a Task?
Swaroop Mishra and Bhavdeep Singh Sachdeva 169

Overview of the SustaiNLP 2020 Shared Task
Alex Wang and Thomas Wolf 174