AACL 2022

SUMEval 2022
Scaling Up Multilingual Evaluation

Proceedings of the Workshop

November 20, 2022

©2022 The Association for Computational Linguistics

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 978-1-952148-25-5

Preface

Massively Multilingual Language Models (MMLMs) are trained on around 100 languages of the world, however, most existing multilingual NLP benchmarks provide evaluation data in only a handful of these languages. The languages present in evaluation benchmarks are usually high-resource and largely belong to the Indo-European language family. This makes current multilingual evaluation unreliable and does not provide a full picture of the performance of MMLMs across the linguistic landscape. Although efforts are being made to create benchmarks that cover a larger variety of tasks, languages, and language families, it is unlikely that we will be able to build benchmarks covering all languages and tasks. Due to this, there is recent interest in alternate strategies for evaluating MMLMs, including performance prediction and Machine Translation of test data. We believe that this is an important yet relatively unexplored area of research that has the potential to make language technologies accessible to all. The SUMEval workshop recieved submissions on techniques for scaling up multilingual evaluation. In addition, the workshop also included a shared task on performance prediction.

Organizing Committee

Kabir Ahuja, Microsoft

Antonios Anastasopoulos, George Mason University

Vishrav Chaudhary, Microsoft

Monojit Choudhury, Microsoft

Sandipan Dandapat, Microsoft

Graham Neubig, Carnegie Mellon University

Barun Patra, Microsoft

Sunayana Sitaram, Microsoft

Table of Contents

The SUMEval 2022 Shared Task on Performance Prediction of Multilingual Pre-trained Language Models
Kabir Ahuja, Antonios Anastasopoulos, Barun Patra, Graham Neubig, Monojit Choudhury, Sandipan Dandapat, Sunayana Sitaram and Vishrav Chaudhary 1

To Train or Not to Train: Predicting the Performance of Massively Multilingual Models
Shantanu Patankar, Omkar Gokhale, Onkar Litake, Aditya Mandke and Dipali Kadam 8

The GMU System Submission for the SUMEval 2022 Shared Task
Syeda Sabrina Akter and Antonios Anastasopoulos 13

NTREX-128 – News Test References for MT Evaluation of 128 Languages
Christian Federmann, Tom Kocmi, Ying Xin 21

IndoRobusta: Towards Robustness Against Diverse Code-Mixed Indonesian Local Languages
Muhammad Farid Adilazuarda, Samuel Cahyawijaya, Genta Indra Winata, Pascale Fung, Ayu Purwarianti 25

Conference Program

November 20, 2022

The SUMEval 2022 Shared Task on Performance Prediction of Multilingual Pre-trained Language Models
Kabir Ahuja, Antonios Anastasopoulos, Barun Patra, Graham Neubig, Monojit Choudhury, Sandipan Dandapat, Sunayana Sitaram and Vishrav Chaudhary

To Train or Not to Train: Predicting the Performance of Massively Multilingual Models
Shantanu Patankar, Omkar Gokhale, Onkar Litake, Aditya Mandke and Dipali Kadam

The GMU System Submission for the SUMEval 2022 Shared Task
Syeda Sabrina Akter and Antonios Anastasopoulos

NTREX-128 – News Test References for MT Evaluation of 128 Languages
Christian Federmann, Tom Kocmi, Ying Xin

IndoRobusta: Towards Robustness Against Diverse Code-Mixed Indonesian Local Languages
Muhammad Farid Adilazuarda, Samuel Cahyawijaya, Genta Indra Winata, Pascale Fung, Ayu Purwarianti