COLING 2025

SUMEval-2
The 2nd Workshop on Scaling Up Multilingual & Multi-Cultural Evaluation

Proceedings of the Workshop

January 20, 2025

©2025 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) like mBERT, XLMR and XY-LENT support around 100 languages of the world. Additionally, generative models like GPT-4 and BLOOM have shown impressive performance in English and a few high-resource languages. However, most existing multilingual NLP benchmarks reflect a handful of cultures and languages. The languages present in evaluation benchmarks are usually high-resource and largely belong to the Indo-European language family. By extension, the cultures represented in evaluation benchmarks are also largely reflective of Western society. This makes current evaluation unreliable and does not provide a full picture of the performance of MMLMs across the linguistic and cultural landscape. Although efforts are being made to create benchmarks that cover a larger variety of tasks, cultures, languages, and language families, it is clear that scaling-up multilingual and multi-cultural evaluation that can eventually lead to better models for all languages and cultures remains a formidable research challenge. This workshop is the second workshop in the SUMEval series, following a successful first workshop SUMEval 2022 co-located with AACL 2022. This year's workshop SUMEval-2, co-located with COLING 2025, has a wider scope focusing on multicultural evaluation in addition to multilingual evaluation.

Organizing Committee

Hellina Hailu Nigatu, UC Berkeley

Monojit Choudhury, MBZUAI

Oana Ignat, Santa Clara University

Sebastian Ruder, Meta

Sunayana Sitaram, Microsoft

Vishrav Chaudhary, Meta

Table of Contents

The First Multilingual Model For The Detection of Suicide Texts
Rodolfo Joel Zevallos, Annika Marie Schoene and John E. Ortega 1

CrossIn: An Efficient Instruction Tuning Approach for Cross-Lingual Knowledge Alignment
Geyu Lin, Bin Wang, Zhengyuan Liu and Nancy F. Chen 12

Evaluating Dialect Robustness of Language Models via Conversation Understanding
Dipankar Srirag, Nihar Ranjan Sahoo and Aditya Joshi 24

Cross-Lingual Document Recommendations with Transformer-Based Representations: Evaluating Multilingual Models and Mapping Techniques
Tsegaye Misikir Tashu, Eduard-Raul Kontos, Matthia Sabatelli and Matias Valdenegro-Toro . . . 39

VRCP: Vocabulary Replacement Continued Pretraining for Efficient Multilingual Language Models
Yuta Nozaki, Dai Nakashima, Ryo Sato and Naoki Asaba 48

Tentative Conference Program

0900 - 0915 Inaugural Remarks

0915 - 1015 Invited Talk by Minsu Park: Cross-Cultural Mood Dynamics and Detection Algorithms

1015 - 1100 Coffee break

1100 - 1200 Panel Discussion 1: Challenges of Collecting Culturally Grounded Multilingual Data for Training and Evaluation of NLP Systems

1200 - 1230 2-min Pitch for Papers (both archival and non-archival papers)

1230 - 1400 Lunch

1400 - 1530 Poster Session (both archival and non-archival)

1530 - 1600 Coffee break

1600 - 1715 Panel Discussion 2: Diverse Cultures. Diverse Problems. Diverse Solutions. Understanding the Nuanced Challenges and Opportunities in Working with Diverse Cultures

1715 - 1730 Closing and Awards