

3rd Workshop on Multi-lingual Representation Learning (MRL 2023)

Singapore
7 December 2023

ISBN: 978-1-7138-8602-0

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© (2023) by the Association for Computational Linguistics
All rights reserved.

Printed with permission by Curran Associates, Inc. (2024)

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

<i>UniBriVL: Robust Audio Representation and Generation of Audio Driven Diffusion Models</i> Sen Fang, Bowen Gao, Yangjian Wu and TeikToe Teoh	1
<i>Meta-learning For Vision-and-language Cross-lingual Transfer</i> Hanxu Hu and Frank Keller	12
<i>Counterfactually Probing Language Identity in Multilingual Models</i> Anirudh Srinivasan, Venkata Subrahmanyam Govindarajan and Kyle Mahowald	24
<i>A General-Purpose Multilingual Document Encoder</i> Onur Galoğlu Robert Litschko, Robert Litschko and Goran Glavaš	37
<i>Zero-Shot Cross-Lingual Sentiment Classification under Distribution Shift: an Exploratory Study</i> Maarten De Raedt, Semere Kiros Bitew, Frédéric Godin, Thomas Demeester and Chris Develder 50	
<i>To token or not to token: A Comparative Study of Text Representations for Cross-Lingual Transfer</i> Md Mushfiqur Rahman, Fardin Ahsan Sakib, Fahim Faisal and Antonios Anastasopoulos	67
<i>Adapt and Prune Strategy for Multilingual Speech Foundational Model on Low-resourced Languages</i> Hyeon Soo Kim, Chung Hyeon Cho, Hyejin Won and Kyung Ho Park	85
<i>Multilingual Word Embeddings for Low-Resource Languages using Anchors and a Chain of Related Languages</i> Viktor Hangya, Silvia Severini, Radoslav Ralev, Alexander Fraser and Hinrich Schütze	95
<i>TalaMT: Multilingual Machine Translation for Cabécar-Bribri-Spanish</i> Alex Jones, Rolando Coto-Solano and Guillermo González Campos	106
<i>Mergen: The First Manchu-Korean Machine Translation Model Trained on Augmented Data</i> Jean Seo, Sungjoo Byun, Minha Kang and Sangah Lee	118
<i>Improving Cross-Lingual Transfer for Open Information Extraction with Linguistic Feature Projection</i> Youmi Ma, Bhushan Kotnis, Carolin Lawrence, Goran Glavaš and Naoaki Okazaki	125
<i>Geographic and Geopolitical Biases of Language Models</i> Fahim Faisal and Antonios Anastasopoulos	139
<i>Task-Based MoE for Multitask Multilingual Machine Translation</i> Hai Pham, Young Jin Kim, Subhabrata Mukherjee, David P. Woodruff, Barnabas Poczos and Hany Hassan	164
<i>Does the English Matter? Elicit Cross-lingual Abilities of Large Language Models</i> Leonardo Ranaldi and Giulia Pucci	173
<i>CAPIVARA: Cost-Efficient Approach for Improving Multilingual CLIP Performance on Low-Resource Languages</i> Gabriel Oliveira dos Santos, Diego Alysson Braga Moreira, Alef Iury Ferreira, Jhessica Silva, Luiz Pereira, Pedro Bueno, Thiago Sousa, Helena Maia, Nádia Da Silva, Esther Colombini, Helio Pedrini and Sandra Avila	184
<i>Code-switching as a cross-lingual Training Signal: an Example with Unsupervised Bilingual Embedding</i> Felix Gaschi, Ilias El-Baamrani, Barbara Gendron, Parisa Rastin and Yannick Toussaint	208

<i>Learning to translate by learning to communicate</i>	
C.M. Downey, Xuhui Zhou, Zeyu Liu and Shane Steinert-Threlkeld	218
<i>Contrastive Learning for Universal Zero-Shot NLI with Cross-Lingual Sentence Embeddings</i>	
Md Kowsher, Md. Shohanur Islam Sobuj, Nusrat Jahan Prottasha, Mohammad Shamsul Arefin and Yasuhiko Morimoto	239
<i>UD-MULTIGENRE – a UD-Based Dataset Enriched with Instance-Level Genre Annotations</i>	
Vera Danilova and Sara Stymne	253
<i>Embedding Structure Matters: Comparing Methods to Adapt Multilingual Vocabularies to New Languages</i>	
C.M. Downey, Terra Blevins, Nora Goldfine and Shane Steinert-Threlkeld	268
<i>Multi-EuP: The Multilingual European Parliament Dataset for Analysis of Bias in Information Retrieval</i>	
Jinrui Yang, Timothy Baldwin and Trevor Cohn	282
<i>Generating Continuations in Multilingual Idiomatic Contexts</i>	
Rhitabrat Pokharel and Ameeta Agrawal	292
<i>CUNI Submission to MRL 2023 Shared Task on Multi-lingual Multi-task Information Retrieval</i>	
Jindřich Helcl and Jindřich Libovický	302
<i>Findings of the 1st Shared Task on Multi-lingual Multi-task Information Retrieval at MRL 2023</i>	
Francesco Tinner, David Ifeoluwa Adelani, Chris Emezue, Mammad Hajili, Omer Goldman, Muhammad Farid Adilazuarda, Muhammad Dehan Al Kautsar, Aziza Mirsaidova, Müge Kural, Dylan Massey, Chiamaka Chukwuneke, Chinedu Mbonu, Damilola Oluwaseun Oloyede, Kayode Olaleye, Jonathan Atala, Benjamin A. Ajibade, Saksham Bassi, Rahul Aralikkatte, Najoung Kim and Duygu Ataman	310