Machine Translation Summit 2023

Volume 1: Research Track

Macau, China
4-8 September 2023

ISBN: 978-1-7138-9040-9

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 authors.
All rights reserved.

These articles are licensed under a Creative Commons Attribution 4.0 International License
(CC BY 4.0).

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

Multiloop Incremental Bootstrapping for Low-Resource Machine Translation
Wuying Liu, Wei Li and Lin Wang 1

Joint Dropout: Improving Generalizability in Low-Resource Neural Machine Translation through Phrase Pair Variables
Ali Araabi, Vlad Niculae and Christof Monz 12

A Study of Multilingual versus Meta-Learning for Language Model Pre-Training for Adaptation to Unseen Low Resource Languages
Jyotsana Khatri, Rudra Murthy, Amar Prakash Azad and Pushpak Bhattacharyya 26

Data Augmentation with Diversified Rephrasing for Low-Resource Neural Machine Translation
Yuan Gao, Feng Hou, Huia Jahnke and Ruili Wang 35

A Dual Reinforcement Method for Data Augmentation using Middle Sentences for Machine Translation
Wenyi TANG and Yves Lepage 48

Perturbation-based QE: An Explainable, Unsupervised Word-level Quality Estimation Method for Black-box Machine Translation
Tu Anh Dinh and Jan Niehues 59

Semi-supervised Learning for Quality Estimation of Machine Translation
Tarun Bhatia, Martin Kraemer, Eduardo Vellasques and Eleftherios Avramidis 72

Learning from Past Mistakes: Quality Estimation from Monolingual Corpora and Machine Translation Learning Stages
Thierry Etchegoyhen and David Ponce 84

Exploring Domain-shared and Domain-specific Knowledge in Multi-Domain Neural Machine Translation
Zhibo Man, YUJIE ZHANG, Yuanmeng Chen, Yufeng Chen and Jinan Xu 99

Enhancing Translation of Myanmar Sign Language by Transfer Learning and Self-Training
Hlaing Myat Nwe, Kiyoaki Shirai, Natthawut Kertkeidkachorn, Thanaruk Theeramunkong, Ye Kyaw Thu, Thepchai Supnithi and Natsuda Kaothanthong 111

Improving Embedding Transfer for Low-Resource Machine Translation
Van Hien Tran, Chenchen Ding, Hideki Tanaka and Masao Utiyama 123

Boosting Unsupervised Machine Translation with Pseudo-Parallel Data
Ivana Kvapilíková and Ondřej Bojar 135

A Study on the Effectiveness of Large Language Models for Translation with Markup
Raj Dabre, Bianka Buschbeck, Miriam Exel and Hideki Tanaka 148

A Case Study on Context Encoding in Multi-Encoder based Document-Level Neural Machine Translation
Ramakrishna Appicharla, Baban Gain, Santanu Pal and Asif Ekbal 160

In-context Learning as Maintaining Coherency: A Study of On-the-fly Machine Translation Using Large Language Models
Suzanna Sia and Kevin Duh 173

Beyond Correlation: Making Sense of the Score Differences of New MT Evaluation Metrics
Chi-kiu Lo, Rebecca Knowles and Cyril Goutte 186

Bad MT Systems are Good for Quality Estimation
Iryna Tryhubyshyn, Aleš Tamchyna and Ondřej Bojar 200

Improving Domain Robustness in Neural Machine Translation with Fused Topic Knowledge Embeddings
Danai Xezonaki, Talaat Khalil, David Stap and Brandon Denis 209

Instance-Based Domain Adaptation for Improving Terminology Translation
Prashanth Nayak, John Kelleher, rejwanul haque and Andy Way 222

Learning from Mistakes: Towards Robust Neural Machine Translation for Disfluent L2 Sentences
Shuyue Stella Li and Philipp Koehn 235

The Role of Compounds in Human vs. Machine Translation Quality
Kristyna Neumannova and Ondřej Bojar 248

Benchmarking Dialectal Arabic-Turkish Machine Translation
Hasan Alkheder, Houda Bouamor, Nizar Habash and Ahmet Zengin 261

Context-aware Neural Machine Translation for English-Japanese Business Scene Dialogues
Sumire Honda, Patrick Fernandes and Chrysoula Zerva 272

A Context-Aware Annotation Framework for Customer Support Live Chat Machine Translation
Miguel Menezes, M. Amin Farajian, Helena Moniz and João Varelas Graça 286

Targeted Data Augmentation Improves Context-aware Neural Machine Translation
Harritxu Gete, Thierry Etchegoyhen and Gorka Labaka 298

Target Language Monolingual Translation Memory based NMT by Cross-lingual Retrieval of Similar Translations and Reranking
Takuya Tamura, Xiaotian Wang, Takehito Utsuro and Masaaki Nagata 313

Towards Zero-Shot Multilingual Poetry Translation
Wai Lei Song, Haoyun Xu, Derek F. Wong, Runzhe Zhan, Lidia S. Chao and Shanshan Wang . 324

Leveraging Highly Accurate Word Alignment for Low Resource Translation by Pretrained Multilingual Model
Jingyi Zhu, Minato Kondo, Takuya Tamura, Takehito Utsuro and Masaaki Nagata 336

Pivot Translation for Zero-resource Language Pairs Based on a Multilingual Pretrained Model
Kenji Imamura, Masao Utiyama and Eiichiro Sumita 348

Character-level NMT and language similarity
Josef Jon and Ondřej Bojar 360

Negative Lexical Constraints in Neural Machine Translation
Josef Jon, Dusan Varis, Michal Novák, João Paulo Aires and Ondřej Bojar 372

Post-editing of Technical Terms based on Bilingual Example Sentences
Elsie K. Y. Chan, John Lee, Chester Cheng and Benjamin Tsou 385

A Filtering Approach to Object Region Detection in Multimodal Machine Translation
Ali Hatami, Paul Buitelaar and Mihael Arcan 393