

Machine Translation Summit XVIII

Volume 1: MT Research Track

Online
16 – 20 August 2021

Editors:

**Kevin Duh
Francisco Guzman
Stephen Richardson**

ISBN: 978-1-7138-3919-4

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© (2021) by the Association for Computational Linguistics
All rights reserved.

Copyright for individual papers remains with the authors and are licensed under a Creative Commons 4.0 license, CC-BY-ND. (<https://creativecommons.org/licenses/by-nd/4.0/>)

Printed with permission by Curran Associates, Inc. (2022)

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

Contents

- 1 Learning Curricula for Multilingual Neural Machine Translation Training
Gaurav Kumar, Philipp Koehn and Sanjeev Khudanpur
- 10 Investigating Active Learning in Interactive Neural Machine Translation
Kamal Gupta, Dhanvanth Boppana, Rejwanul Haque, Asif Ekbal and Pushpak Bhattacharyya
- 23 Crosslingual Embeddings are Essential in UNMT for distant languages: An English to IndoAryan Case Study
Tamali Banerjee, Rudra V Murthy and Pushpak Bhattacharya
- 35 Neural Machine Translation in Low-Resource Setting: a Case Study in English-Marathi Pair
Aakash Banerjee, Aditya Jain, Shivam Mhaskar, Sourabh Dattatray Deoghare, Aman Sehgal and Pushpak Bhattacharya
- 48 Transformers for Low-Resource Languages: Is Féidir Linn!
Seamus Lankford, Haithem Alfi and Andy Way
- 61 The Effect of Domain and Diacritics in Yoruba--English Neural Machine Translation
David Adelani, Dana Ruitter, Jesujoba Alabi, Damilola Adebajo, Adesina Ayeni, Mofe Adeyemi, Ayodele Esther Awokoya and Cristina España-Bonet
- 76 Integrating Unsupervised Data Generation into Self-Supervised Neural Machine Translation for Low-Resource Languages
Dana Ruitter, Dietrich Klakow, Josef van Genabith and Cristina España-Bonet

- 92 Surprise Language Challenge: Developing a Neural Machine Translation System between Pashto and English in Two Months
Alexandra Birch, Barry Haddow, Antonio Valerio Miceli Barone, Jindrich Helcl, Jonas Waldendorf, Felipe Sánchez Martínez, Mikel Forcada, Víctor Sánchez Cartagena, Juan Antonio Pérez-Ortiz, Miquel Esplà-Gomis, Wilker Aziz, Lina Murady, Sevi Sariisik, Peggy van der Kreeft and Kay Macquarrie
- 103 Like Chalk and Cheese? On the Effects of Translationese in MT Training
Samuel Larkin, Michel Simard and Rebecca Knowles
- 114 Investigating Softmax Tempering for Training Neural Machine Translation Models
Raj Dabre and Atsushi Fujita
- 127 Scrambled Translation Problem: A Problem of Denoising UNMT
Tamali Banerjee, Rudra V Murthy and Pushpak Bhattacharya
- 139 Make the Blind Translator See The World: A Novel Transfer Learning Solution for Multimodal Machine Translation
Minghan Wang, Jiaxin Guo, Yimeng Chen, Chang Su, Min Zhang, Shimin Tao and Hao Yang
- 150 Sentiment Preservation in Review Translation using Curriculum-based Re-inforcement Framework
Divya Kumari, Soumya Chennabasavaraj, Nikesh Garera and Asif Ekbal
- 163 On nature and causes of observed MT errors
Maja Popovic
- 176 A Comparison of Sentence-Weighting Techniques for NMT
Simon Rieß, Matthias Huck and Alex Fraser

- 188 Sentiment-based Candidate Selection for NMT
Alexander G Jones and Derry Wijaya
- 202 Studying The Impact Of Document-level Context On Simultaneous Neural Machine Translation
Raj Dabre, Aizhan Imankulova and Masahiro Kaneko
- 215 Attainable Text-to-Text Machine Translation vs. Translation: Issues Beyond Linguistic Processing
Atsushi Fujita
- 231 Modeling Target-side Inflection in Placeholder Translation
Ryokan Ri, Toshiaki Nakazawa and Yoshimasa Tsuruoka
- 243 Product Review Translation using Phrase Replacement and Attention Guided Noise Augmentation
Kamal Gupta, Soumya Chennabasavaraj, Nikesh Garera and Asif Ekbal
- 256 Optimizing Word Alignments with Better Subword Tokenization
Anh Khoa Ngo Ho and François Yvon
- 270 Introducing Mouse Actions into Interactive-Predictive Neural Machine Translation
Ángel Navarro and Francisco Casacuberta
- 282 Neural Machine Translation with Inflected Lexicon
Artur Nowakowski and Krzysztof Jassem
- 293 An Alignment-Based Approach to Semi-Supervised Bilingual Lexicon Induction with Small Parallel Corpora
Kelly V Marchisio, Philipp Koehn and Conghao Xiong