MT Summit 2023

MTS Machine Translation Summit 2023

September 4-8, 2023 Macau SAR, China

Proceedings of Machine Translation Summit XIX
Vol. 1: Research Track

September 4 - 8, 2023

Introduction

The research track at MT Summit 2023 has a wide range of topics with 33 papers selected from entire 50 submissions. The part of subjects covered by the research track, as indicated by the keywords in the titles below:

- Low-Resource, Zero-resource MT
- Document-Level, Coherent, Context-aware NMT
- Quality Estimation
- Multi-domain, Domain Robustness, Domain Adaptation
- Unsupervised NMT
- Robust NMT, Markup Translation
- MT Evaluation
- Annotation
- Poetry, Compounds, Dialectal
- Post-editing
- Sign Language, Multimodal

Among the 33 papers, 19 papers are accepted as oral presentations and 14 as poster presentations. The most popular subject is "Low-Resource" MT. The subjects of "Context-aware" NMT and "Quality Estimation" are also popular. We also have unique topics like Myanmar Sign Language, Translation with Markup, Robust NMT, Dialectal Arabic-Turkish MT, and Poetry Translation. These indicate we have both popular topics and unique topics, which could be overlooked in the larger general NLP conferences.

We thank the authors, reviewers, and MT Summit organizing committee for making a good conference happen. We also thank our invited speakers for the research track for sharing their interesting experiences: Min Zhang, Ondřej Bojar, Mitesh Khapra, Tong Xiao, and Isao Goto.

Sincerely,
Masao Utiyama and Rui Wang (Research Track Co-Chairs)

Organizing Committee

General Chair

Eiichiro Sumita, National Institute of Information and Communications Technology

Steering Committee

Eiichiro Sumita, National Institute of Information and Communications Technology
Kozo Moriguchi, Kawamura International Co. Ltd.
Derek Wong, University of Macau
Sadao Kurohashi, National Institute of Informatics & Kyoto University
Hideki Tanaka, National Institute of Information and Communications Technology

Research Track Chair

Masao Utiyama, National Institute of Information and Communications Technology
Rui Wang, Shanghai Jiao Tong University

Users Track Chair

Masaru Yamada, Rikkyo University
Félix do Carmo, University of Surrey

Workshop Chair

Jiajun Zhang, Chinese Academy of Sciences
Thepchai Supnithi, The National Electronics and Computer Technology Center

Local Arrangement Chair

Derek Wong, University of Macau
Hou Pong Chan, University of Macau

Publication Chair

Katsuhito Sudoh, Nara Institute of Science and Technology
Xuebo Liu, Harbin Institute of Technology, Shenzhen

Sponsorship Chair

Kozo Moriguchi, Kawamura International Co. Ltd.
Jaap van der Meer, TAUS
Tong Xiao, Northeastern University

Conference Manager

Andrew Jiang, Macau Expo Group

Program Committee

Sweta Agrawal
Loic Barrault
Pushpak Bhattacharyya
Houda Bouamor
Michael Carl
Abhisek Chakrabarty
Xingyu Chen
Colin Cherry
Chenhui Chu
Chenchen Ding
Kevin Duh
Marcello Federico
Christian Federmann
Minwei Feng
Markus Freitag
Ruize Gao
Isao Goto
Barry Haddow
Zhiwei He
MukendiI Kalamba
Rebecca Knowles
Philipp Koehn
Mamoru Komachi
Shankar Kumar
Lemao Liu
Qun Liu
Wolfgang Macherey
Arne Mauser
Arul Menezes
Hideya Mino
Makoto Morishita
Masaaki Nagata
Hiromi Nakaiwa
Toshiaki Nakazawa
Takashi Ninomiya
Shantipriya Parida
Tharindu Ranasinghe
Raphael Rubino
Fatiha Sadat
Michel Simard
Haiyue Song
Katsuhito Sudoh
Akihiro Tamura
Hajime Tsukada
Yiming Wang
Xiaolin Wang
Taro Watanabe
Muyun Yang
François Yvon
Rabih Zbib
Jiajun Zhang
Pei Zhang
Bing Zhao

Table of Contents

Conference Program

Wednesday, 6th September

11:15–12:15 Session RP1: Research Track Posters (1)

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

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

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

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

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

16:00–17:30 Session RS1: Quality Estimation

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

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

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

Thursday, 7th September

10:30–12:00 Session RS2: Transfer Learning Approach

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

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

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

10:30–12:00 Session RS3: Training with Auxiliary Information

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

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

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

Thursday, 7th September (continued)

15:00–16:00 Session RP2: Research Track Posters (2)

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

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

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

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

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

16:00–17:30 Session RS4: Feedback and Evaluation

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

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

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

Friday, 8th September

10:30–12:00 Session RS5: Context-aware Machine Translation

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

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

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

14:00–16:00 Session RS6: Multilingual Machine Translation

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

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

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

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

Friday, 8th September (continued)

16:00–17:00 Session RP3: Research Track Posters (3)

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

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

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

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