

WAT 2019

The 6th Workshop on Asian Translation

Proceedings of the 6th Workshop on Asian Translation

November 4, 2019
Hong Kong, China

©2019 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 978-1-950737-87-1

Preface

Many Asian countries are rapidly growing these days and the importance of communicating and exchanging the information with these countries has intensified. To satisfy the demand for communication among these countries, machine translation technology is essential.

Machine translation technology has rapidly evolved recently and it is seeing practical use especially between European languages. However, the translation quality of Asian languages is not that high compared to that of European languages, and machine translation technology for these languages has not reached a stage of proliferation yet. This is not only due to the lack of the language resources for Asian languages but also due to the lack of techniques to correctly transfer the meaning of sentences from/to Asian languages. Consequently, a place for gathering and sharing the resources and knowledge about Asian language translation is necessary to enhance machine translation research for Asian languages.

The Workshop on Machine Translation (WMT), the world's largest machine translation workshop, mainly targets on European languages and does not include Asian languages. The International Workshop on Spoken Language Translation (IWSLT) has spoken language translation tasks for some Asian languages using TED talk data, but there is no task for written language.

The Workshop on Asian Translation (WAT) is an open machine translation evaluation campaign focusing on Asian languages. WAT gathers and shares the resources and knowledge of Asian language translation to understand the problems to be solved for the practical use of machine translation technologies among all Asian countries. WAT is unique in that it is an "open innovation platform": the test data is fixed and open, so participants can repeat evaluations on the same data and confirm changes in translation accuracy over time. WAT has no deadline for the automatic translation quality evaluation (continuous evaluation), so participants can submit translation results at any time.

Following the success of the previous WAT workshops (WAT2014 – WAT2018), WAT2019 will bring together machine translation researchers and users to try, evaluate, share and discuss brand-new ideas about machine translation. For the 6th WAT, we included 5 new translation subtasks. We had 25 teams who submitted their translation results, and about 400 submissions in total.

In addition to the shared tasks, WAT2019 also features scientific papers on topics related to the machine translation, especially for Asian languages. The program committee accepted 6 papers, which focus on neural machine translation, and construction and evaluation of language resources.

We are grateful to "SunFlare Co., Ltd.", "Japan Exchange Group, Inc. (JPX)", "Asia-Pacific Association for Machine Translation (AAMT)" and "Kawamura International" for partially sponsoring the workshop. We would like to thank all the authors who submitted papers. We express our deepest gratitude to the committee members for their timely reviews. We also thank the EMNLP-IJCNLP 2019 organizers for their help with administrative matters.

WAT 2019 Organizers

Organizers:

Toshiaki Nakazawa, The University of Tokyo, Japan
Chenchen Ding, National Institute of Information and Communications Technology (NICT), Japan
Raj Dabre, National Institute of Information and Communications Technology (NICT), Japan
Anoop Kunchukuttan, Microsoft AI and Research, India
Win Pa Pa, University of Computer Studies, Yangon (UCSY), Myanmar
Nobushige Doi, Japan Exchange Group (JPX), Japan
Yusuke Oda, Google, Japan
Ondřej Bojar, Charles University, Prague, Czech Republic
Shantipriya Parida, Idiap Research Institute, Martigny, Switzerland
Isao Goto, Japan Broadcasting Corporation (NHK), Japan
Hideya Mino, Japan Broadcasting Corporation (NHK), Japan
Hiroshi Manabe, National Institute of Information and Communications Technology (NICT), Japan
Katsuhito Sudoh, Nara Institute of Science and Technology (NAIST), Japan
Sadao Kurohashi, Kyoto University, Japan
Pushpak Bhattacharyya, Indian Institute of Technology Bombay (IIT), India

Program Committee:

Hailong Cao, Harbin Institute of Technology, China
Michael Carl, Kent State University, USA
Chenhui Chu, Osaka University, Japan
Fabien Cromières, Kyoto University, Japan
Kenji Imamura, NICT, Japan
Hideto Kazawa, Google, Japan
Anoop Kunchookuttan, Microsoft, India
Yang Liu, Tsinghua University, China
Takashi Ninomiya, Ehime University, Japan
Liling Tan, Rakuten Institute of Technology, Singapore
Masao Utiyama, NICT, Japan
Jiajun Zhang, Chinese Academy of Sciences, China
Dongdong Zhang, Microsoft Research Asia, China

Technical Collaborators:

Luis Fernando D'Haro, Universidad Politécnica de Madrid, Spain

Rafael E. Banchs, Nanyang Technological University, Singapore

Haizhou Li, National University of Singapore, Singapore

Invited talk: Multitask Learning from Multilingual Multimodal Data

Desmond Elliott

The University of Copenhagen

Abstract

I will talk about two perspectives on learning from multilingual multimodal data: as a language generation problem and as cross-modal retrieval problem. In the language generation problem of multimodal machine translation, I will discuss whether we should learn grounded representations by using the additional visual context as a conditioning input or as a variable that the model learns to predict, and highlight some recent arguments about whether models are actually sensitive to the visual context. As a multilingual image–sentence retrieval problem, I will discuss experiments that highlight situations in which it is useful to train with multilingual annotations, as opposed to monolingual annotations, and the challenges in learning from disjoint cross-lingual datasets.

Biography

Desmond is an Assistant Professor at the University of Copenhagen. He received his PhD from the University of Edinburgh, and was a postdoctoral researcher at CWI Amsterdam, the University of Amsterdam, and the University of Edinburgh, funded by an Alain Bensoussan Career Development Fellowship and an Amazon Research Award. His research interests include multimodal and multilingual machine learning, which has appeared in papers ACL, CoNLL, EMNLP and NAACL. He was involved in the creation the Multi30K and How2 multilingual multimodal datasets and has developed a variety of models that learn from these types of data. He co-organised the How 2 Challenge Workshop at ICML 2019, the Multimodal Machine Translation Shared Task from 2016–2018, and the 2018 Frederick Jelinek Memorial Workshop on Grounded Sequence-to-Sequence Learning.

Table of Contents

<i>Overview of the 6th Workshop on Asian Translation</i> Toshiaki Nakazawa, Nobushige Doi, Shohei Higashiyama, Chenchen Ding, Raj Dabre, Hideya Mino, Isao Goto, Win Pa Pa, Anoop Kunchukuttan, Shantipriya Parida, Ondřej Bojar and Sadao Kurohashi	1
<i>Compact and Robust Models for Japanese-English Character-level Machine Translation</i> Jinan Dai and Kazunori Yamaguchi	36
<i>Controlling Japanese Honorifics in English-to-Japanese Neural Machine Translation</i> Weston Feely, Eva Hasler and Adrià de Gispert	45
<i>Designing the Business Conversation Corpus</i> Matīss Rikters, Ryokan Ri, Tong Li and Toshiaki Nakazawa	54
<i>English to Hindi Multi-modal Neural Machine Translation and Hindi Image Captioning</i> Sahinur Rahman Laskar, Rohit Pratap Singh, Partha Pakray and Sivaji Bandyopadhyay	62
<i>Supervised and Unsupervised Machine Translation for Myanmar-English and Khmer-English</i> Benjamin Marie, Hour Kaing, Aye Myat Mon, Chenchen Ding, Atsushi Fujita, Masao Utiyama and Eiichiro Sumita	68
<i>NICT's participation to WAT 2019: Multilingualism and Multi-step Fine-Tuning for Low Resource NMT</i> Raj Dabre and Eiichiro Sumita	76
<i>KNU-HYUNDAI's NMT system for Scientific Paper and Patent Tasks onWAT 2019</i> Cheoneum Park, Young-Jun Jung, Kihoon Kim, Geonyeong Kim, Jae-Won Jeon, Seongmin Lee, Junseok Kim and Changki Lee	81
<i>English-Myanmar Supervised and Unsupervised NMT: NICT's Machine Translation Systems at WAT-2019</i> Rui Wang, Haipeng Sun, Kehai Chen, Chenchen Ding, Masao Utiyama and Eiichiro Sumita ...	90
<i>UCSMNLP: Statistical Machine Translation for WAT 2019</i> Aye Thida, Nway Nway Han, Sheinn Thawtar Oo and Khin Thet Htar	94
<i>NTT Neural Machine Translation Systems at WAT 2019</i> Makoto Morishita, Jun Suzuki and Masaaki Nagata	99
<i>Neural Machine Translation System using a Content-equivalently Translated Parallel Corpus for the Newswire Translation Tasks at WAT 2019</i> Hideya Mino, Hitoshi Ito, Isao Goto, Ichiro Yamada, Hideki Tanaka and Takenobu Tokunaga .	106
<i>Facebook AI's WAT19 Myanmar-English Translation Task Submission</i> Peng-Jen Chen, Jiajun Shen, Matthew Le, Vishrav Chaudhary, Ahmed El-Kishky, Guillaume Wenzek, Myle Ott and Marc' Aurelio Ranzato	112
<i>Combining Translation Memory with Neural Machine Translation</i> Akiko Eriguchi, Spencer Rarrick and Hitokazu Matsushita	123
<i>CVIT's submissions to WAT-2019</i> Jerin Philip, Shashank Siripragada, Upendra Kumar, Vinay Namboodiri and C V Jawahar	131

<i>LTRC-MT Simple & Effective Hindi-English Neural Machine Translation Systems at WAT 2019</i> Vikrant Goyal and Dipti Misra Sharma	137
<i>Long Warm-up and Self-Training: Training Strategies of NICT-2 NMT System at WAT-2019</i> Kenji Imamura and Eiichiro Sumita	141
<i>Supervised neural machine translation based on data augmentation and improved training & inference process</i> Yixuan Tong, Liang Liang, Boyan Liu, Shanshan Jiang and Bin Dong.....	147
<i>Sarah’s Participation in WAT 2019</i> Raymond Hendy Susanto, Ohnmar Htun and Liling Tan	152
<i>Our Neural Machine Translation Systems for WAT 2019</i> Wei Yang and Jun Ogata	159
<i>Japanese-Russian TMU Neural Machine Translation System using Multilingual Model for WAT 2019</i> Aizhan Imankulova, Masahiro Kaneko and Mamoru Komachi.....	165
<i>NLPRL at WAT2019: Transformer-based Tamil – English Indic Task Neural Machine Translation System</i> Amit Kumar and Anil Kumar Singh	171
<i>Idiap NMT System for WAT 2019 Multimodal Translation Task</i> Shantipriya Parida, Ondřej Bojar and Petr Motlicek	175
<i>WAT2019: English-Hindi Translation on Hindi Visual Genome Dataset</i> Loitongbam Sanayai Meetei, Thoudam Doren Singh and Sivaji Bandyopadhyay	181
<i>SYSTRAN @ WAT 2019: Russian-Japanese News Commentary task</i> Jitao Xu, TuAnh Nguyen, MinhQuang PHAM, Josep Crego and Jean Senellart	189
<i>UCSYNLP-Lab Machine Translation Systems for WAT 2019</i> Yimon ShweSin, Win Pa Pa and KhinMar Soe.....	195
<i>Sentiment Aware Neural Machine Translation</i> Chenglei Si, Kui Wu, Ai Ti Aw and Min-Yen Kan.....	200
<i>Overcoming the Rare Word Problem for low-resource language pairs in Neural Machine Translation</i> Thi-Vinh Ngo, Thanh-Le Ha, Phuong-Thai Nguyen and Le-Minh Nguyen.....	207
<i>Neural Arabic Text Diacritization: State of the Art Results and a Novel Approach for Machine Translation</i> Ali Fadel, Ibraheem Tuffaha, Bara’ Al-Jawarneh and Mahmoud Al-Ayyoub	215

Workshop Program

Monday, November 4, 2019

09:00–09:30 Welcome and Overview

Overview of the 6th Workshop on Asian Translation

Toshiaki Nakazawa, Nobushige Doi, Shohei Higashiyama, Chenchen Ding, Raj Dabre, Hideya Mino, Isao Goto, Win Pa Pa, Anoop Kunchukuttan, Shantipriya Parida, Ondřej Bojar and Sadao Kurohashi

09:30–10:30 Research Paper I

Compact and Robust Models for Japanese-English Character-level Machine Translation

Jinan Dai and Kazunori Yamaguchi

Controlling Japanese Honorifics in English-to-Japanese Neural Machine Translation

Weston Feely, Eva Hasler and Adrià de Gispert

Designing the Business Conversation Corpus

Matiss Rikters, Ryokan Ri, Tong Li and Toshiaki Nakazawa

10:30–10:50 Poster Booster I

10:50–12:30 System Description Paper (Poster) I

English to Hindi Multi-modal Neural Machine Translation and Hindi Image Captioning

Sahinur Rahman Laskar, Rohit Pratap Singh, Partha Pakray and Sivaji Bandyopadhyay

Supervised and Unsupervised Machine Translation for Myanmar-English and Khmer-English

Benjamin Marie, Hour Kaing, Aye Myat Mon, Chenchen Ding, Atsushi Fujita, Masao Utiyama and Eiichiro Sumita

NICT's participation to WAT 2019: Multilingualism and Multi-step Fine-Tuning for Low Resource NMT

Raj Dabre and Eiichiro Sumita

KNU-HYUNDAI's NMT system for Scientific Paper and Patent Tasks on WAT 2019

Cheoneum Park, Young-Jun Jung, Kihoon Kim, Geonyeong Kim, Jae-Won Jeon, Seongmin Lee, Junseok Kim and Changki Lee

Monday, November 4, 2019 (continued)

English-Myanmar Supervised and Unsupervised NMT: NICT's Machine Translation Systems at WAT-2019

Rui Wang, Haipeng Sun, Kehai Chen, Chenchen Ding, Masao Utiyama and Eiichiro Sumita

UCSMNLP: Statistical Machine Translation for WAT 2019

Aye Thida, Nway Nway Han, Sheinn Thawtar Oo and Khin Thet Htar

NTT Neural Machine Translation Systems at WAT 2019

Makoto Morishita, Jun Suzuki and Masaaki Nagata

Neural Machine Translation System using a Content-equivalently Translated Parallel Corpus for the Newswire Translation Tasks at WAT 2019

Hideya Mino, Hitoshi Ito, Isao Goto, Ichiro Yamada, Hideki Tanaka and Takenobu Tokunaga

Facebook AI's WAT19 Myanmar-English Translation Task Submission

Peng-Jen Chen, Jiajun Shen, Matthew Le, Vishrav Chaudhary, Ahmed El-Kishky, Guillaume Wenzek, Myle Ott and Marc' Aurelio Ranzato

Combining Translation Memory with Neural Machine Translation

Akiko Eriguchi, Spencer Rarrick and Hitokazu Matsushita

CVIT's submissions to WAT-2019

Jerin Philip, Shashank Siripragada, Upendra Kumar, Vinay Namboodiri and C V Jawahar

LTRC-MT Simple & Effective Hindi-English Neural Machine Translation Systems at WAT 2019

Vikrant Goyal and Dipti Misra Sharma

Monday, November 4, 2019 (continued)

12:30–14:00 Lunch Break

14:00–14:45 Invited Talk by Dr. Desmond Elliott

14:45–15:05 Poster Booster II

15:05–15:10 Commemorative Photo

15:10–16:30 System Description Paper (Poster) II

Long Warm-up and Self-Training: Training Strategies of NICT-2 NMT System at WAT-2019

Kenji Imamura and Eiichiro Sumita

Supervised neural machine translation based on data augmentation and improved training & inference process

Yixuan Tong, Liang Liang, Boyan Liu, Shanshan Jiang and Bin Dong

Sarah's Participation in WAT 2019

Raymond Hendy Susanto, Ohnmar Htun and Liling Tan

Our Neural Machine Translation Systems for WAT 2019

Wei Yang and Jun Ogata

Japanese-Russian TMU Neural Machine Translation System using Multilingual Model for WAT 2019

Aizhan Imankulova, Masahiro Kaneko and Mamoru Komachi

NLPRL at WAT2019: Transformer-based Tamil – English Indic Task Neural Machine Translation System

Amit Kumar and Anil Kumar Singh

Idiap NMT System for WAT 2019 Multimodal Translation Task

Shantipriya Parida, Ondřej Bojar and Petr Motlicek

WAT2019: English-Hindi Translation on Hindi Visual Genome Dataset

Loitongbam Sanayai Meetei, Thoudam Doren Singh and Sivaji Bandyopadhyay

Monday, November 4, 2019 (continued)

SYSTRAN @ WAT 2019: Russian-Japanese News Commentary task

Jitao Xu, TuAnh Nguyen, MinhQuang PHAM, Josep Crego and Jean Senellart

UCSYNLP-Lab Machine Translation Systems for WAT 2019

Yimon ShweSin, Win Pa Pa and KhinMar Soe

16:30–17:30 Research Paper II

Sentiment Aware Neural Machine Translation

Chenglei Si, Kui Wu, Ai Ti Aw and Min-Yen Kan

Overcoming the Rare Word Problem for low-resource language pairs in Neural Machine Translation

Thi-Vinh Ngo, Thanh-Le Ha, Phuong-Thai Nguyen and Le-Minh Nguyen

Neural Arabic Text Diacritization: State of the Art Results and a Novel Approach for Machine Translation

Ali Fadel, Ibraheem Tuffaha, Bara' Al-Jawarneh and Mahmoud Al-Ayyoub

17:30–17:35 Closing