WAT 2020

The 7th Workshop on Asian Translation

Proceedings of the Workshop

December 4, 2020

©2020 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 978-1-952148-95-8

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 Conference 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 – WAT2019), WAT2020 will bring together machine translation researchers and users to try, evaluate, share and discuss brand-new ideas about machine translation. For the 7th WAT, we included 6 new translation subtasks. We had 20 teams participated in the shared tasks and 14 teams submitted their translation results for the human evaluation. About 500 translation results were submitted to the automatic evaluation server, and selected submissions were manually evaluated. In addition to the shared tasks, WAT2020 also features scientific papers on topics related to 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." and "Asia-Pacific Association for Machine Translation (AAMT)" 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 AACL-IJCNLP 2020 organizers for their help with administrative matters.

WAT 2020 Organizers

Organizers:

Toshiaki Nakazawa, The University of Tokyo, Japan

Hideki Nakayama, 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

Shohei Higashiyama, 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

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:

Michael Carl, Kent State University, USA

Chenhui Chu, Osaka University, Japan

Fabien Cromières, Free, France

Hideto Kazawa, Google, Japan

Anoop Kunchookuttan, Microsoft AI and Research, India

Kenji Imamura, NICT, Japan

Liling Tan, Rakuten Institute of Technology, Singapore

Masao Utiyama, NICT, Japan

Yang Liu, Tsinghua University, China

Takashi Ninomiya, Ehime University, Japan

Jiajun Zhang, Chinese Academy of Sciences, China

Gurpreet Singh Lehal, Punjabi University, Patiala, India

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

Table of Contents

Overview of the 7th Workshop on Asian Translation
Toshiaki Nakazawa, Hideki Nakayama, Chenchen Ding, Raj Dabre, Shohei Higashiyama, Hideya Mino, Isao Goto, Win Pa Pa, Anoop Kunchukuttan, Shantipriya Parida, Ondřej Bojar and Sadao Kurohashi 1

An Effective Optimization Method for Neural Machine Translation: The Case of English-Persian Bilingually Low-Resource Scenario
Benyamin Ahmadnia and Bonnie Dorr 45

Transformer-based Double-token Bidirectional Autoregressive Decoding in Neural Machine Translation
Kenji Imamura and Eiichiro Sumita 50

Translation of New Named Entities from English to Chinese
Zizheng Zhang, Tosho Hirasawa, Wei Houjing, Masahiro Kaneko and Mamoru Komachi 58

Meta Ensemble for Japanese-Chinese Neural Machine Translation: Kyoto-U+ECNU Participation to WAT 2020
Zhuoyuan Mao, Yibin Shen, Chenhui Chu, Sadao Kurohashi and Cheqing Jin 64

Neural Machine Translation Using Extracted Context Based on Deep Analysis for the Japanese-English Newswire Task at WAT 2020
Isao Goto, Hideya Mino, Hitoshi Ito, Kazutaka Kinugawa, Ichiro Yamada and Hideki Tanaka . . 72

TMU Japanese-English Multimodal Machine Translation System for WAT 2020
Hiroto Tamura, Tosho Hirasawa, Masahiro Kaneko and Mamoru Komachi 80

HW-TSC's Participation in the WAT 2020 Indic Languages Multilingual Task
Zhengzhe Yu, Zhanglin Wu, Xiaoyu Chen, Daimeng Wei, Hengchao Shang, Jiaxin Guo, Zongyao Li, Minghan Wang, Liangyou Li, Lizhi Lei, Hao Yang and Ying Qin 92

NICT's Submission To WAT 2020: How Effective Are Simple Many-To-Many Neural Machine Translation Models?
Raj Dabre and Abhisek Chakrabarty 98

ODIANLP's Participation in WAT2020
Shantipriya Parida, Petr Motlicek, Amulya Ratna Dash, Satya Ranjan Dash, Debasish Kumar Mallick, Satya Prakash Biswal, Priyanka Pattnaik, Biranchi Narayan Nayak and Ondřej Bojar 103

Multimodal Neural Machine Translation for English to Hindi
Sahinur Rahman Laskar, Abdullah Faiz Ur Rahman Khilji, Partha Pakray and Sivaji Bandyopadhyay 109

The ADAPT Centre's Participation in WAT 2020 English-to-Odia Translation Task
Prashanth Nayak, Rejwanul Haque and Andy Way 114

NLPRL Odia-English: Indic Language Neural Machine Translation System
Rupjyoti Baruah and Rajesh Kumar Mundotiya 118

WT: Wipro AI Submissions to the WAT 2020
Santanu Pal 122

Korean-to-Japanese Neural Machine Translation System using Hanja Information
Hwichan Kim, Tosho Hirasawa and Mamoru Komachi 127

Goku's Participation in WAT 2020
Dongzhe Wang and Ohnmar Htun 135

The ADAPT Centre's Neural MT Systems for the WAT 2020 Document-Level Translation Task
Wandri Jooste, Rejwanul Haque and Andy Way 142

The University of Tokyo's Submissions to the WAT 2020 Shared Task
Matīss Rikters, Toshiaki Nakazawa and Ryokan Ri 147

Improving NMT via Filtered Back Translation
Nikhil Jaiswal, Mayur Patidar, Surabhi Kumari, Manasi Patwardhan, Shirish Karande, Puneet Agarwal and Lovekesh Vig 154

A Parallel Evaluation Data Set of Software Documentation with Document Structure Annotation
Bianka Buschbeck and Miriam Exel 160

Inference-only sub-character decomposition improves translation of unseen logographic characters
Danielle Saunders, Weston Feely and Bill Byrne 170

An Error-based Investigation of Statistical and Neural Machine Translation Performance on Hindi-to-Tamil and English-to-Tamil
Akshai Ramesh, Venkatesh Balavadhani Parthasa, Rejwanul Haque and Andy Way 178

Workshop Program

Friday, December 4, 2020

10:00–10:05 Welcome

Overview of the 7th Workshop on Asian Translation
Toshiaki Nakazawa, Hideki Nakayama, Chenchen Ding, Raj Dabre, Shohei Higashiyama, Hideya Mino, Isao Goto, Win Pa Pa, Anoop Kunchukuttan, Shantipriya Parida, Ondřej Bojar and Sadao Kurohashi

10:05–11:05 Research Paper I

An Effective Optimization Method for Neural Machine Translation: The Case of English-Persian Bilingually Low-Resource Scenario
Benyamin Ahmadnia and Bonnie Dorr

Transformer-based Double-token Bidirectional Autoregressive Decoding in Neural Machine Translation
Kenji Imamura and Eiichiro Sumita

Translation of New Named Entities from English to Chinese
Zizheng Zhang, Tosho Hirasawa, Wei Houjing, Masahiro Kaneko and Mamoru Komachi

11:05–11:10 Short Break

11:10–12:00 Shared Task I - ASPEC / JIJI / Multimodal En-Ja

Task Descriptions and Results
Toshiaki Nakazawa, Isao Goto and Hideki Nakayama

Meta Ensemble for Japanese-Chinese Neural Machine Translation: Kyoto-U+ECNU Participation to WAT 2020
Zhuoyuan Mao, Yibin Shen, Chenhui Chu, Sadao Kurohashi and Cheqing Jin

Neural Machine Translation Using Extracted Context Based on Deep Analysis for the Japanese-English Newswire Task at WAT 2020
Isao Goto, Hideya Mino, Hitoshi Ito, Kazutaka Kinugawa, Ichiro Yamada and Hideki Tanaka

TMU Japanese-English Multimodal Machine Translation System for WAT 2020
Hiroto Tamura, Tosho Hirasawa, Masahiro Kaneko and Mamoru Komachi

Friday, December 4, 2020 (continued)

12:00–14:00 Lunch Break

14:05–15:39 Shared Task II - Indic Multilingual / UFAL EnOdia / Multimodal En-Hi / Hinden

Task Descriptions and Results
Raj Dabre and Shantipriya Parida

HW-TSC's Participation in the WAT 2020 Indic Languages Multilingual Task
Zhengzhe Yu, Zhanglin Wu, Xiaoyu Chen, Daimeng Wei, Hengchao Shang, Jiaxin Guo, Zongyao Li, Minghan Wang, Liangyou Li, Lizhi Lei, Hao Yang and Ying Qin

NICT's Submission To WAT 2020: How Effective Are Simple Many-To-Many Neural Machine Translation Models?
Raj Dabre and Abhisek Chakrabarty

ODIANLP's Participation in WAT2020
Shantipriya Parida, Petr Motlicek, Amulya Ratna Dash, Satya Ranjan Dash, Debasish Kumar Mallick, Satya Prakash Biswal, Priyanka Pattnaik, Biranchi Narayan Nayak and Ondřej Bojar

Multimodal Neural Machine Translation for English to Hindi
Sahinur Rahman Laskar, Abdullah Faiz Ur Rahman Khilji, Partha Pakray and Sivaji Bandyopadhyay

The ADAPT Centre's Participation in WAT 2020 English-to-Odia Translation Task
Prashanth Nayak, Rejwanul Haque and Andy Way

NLPRL Odia-English: Indic Language Neural Machine Translation System
Rupjyoti Baruah and Rajesh Kumar Mundotiya

WT: Wipro AI Submissions to the WAT 2020
Santanu Pal

Friday, December 4, 2020 (continued)

15:39–15:45 Short Break

15:45–16:55 Shared Task III - JPC / BSD

Task Descriptions and Results
Chenchen Ding and Toshiaki Nakazawa

Korean-to-Japanese Neural Machine Translation System using Hanja Information
Hwichan Kim, Tosho Hirasawa and Mamoru Komachi

Goku's Participation in WAT 2020
Dongzhe Wang and Ohnmar Htun

The ADAPT Centre's Neural MT Systems for the WAT 2020 Document-Level Translation Task
Wandri Jooste, Rejwanul Haque and Andy Way

The University of Tokyo's Submissions to the WAT 2020 Shared Task
Matīss Rikters, Toshiaki Nakazawa and Ryokan Ri

Improving NMT via Filtered Back Translation
Nikhil Jaiswal, Mayur Patidar, Surabhi Kumari, Manasi Patwardhan, Shirish Karande, Puneet Agarwal and Lovekesh Vig

16:55–17:00 Short Break

Friday, December 4, 2020 (continued)

17:00–18:00 Research Paper II

A Parallel Evaluation Data Set of Software Documentation with Document Structure Annotation
Bianka Buschbeck and Miriam Exel

Inference-only sub-character decomposition improves translation of unseen logographic characters
Danielle Saunders, Weston Feely and Bill Byrne

An Error-based Investigation of Statistical and Neural Machine Translation Performance on Hindi-to-Tamil and English-to-Tamil
Akshai Ramesh, Venkatesh Balavadhani Parthasa, Rejwanul Haque and Andy Way

18:00–18:05 Closing