COLING

International Conference on Computational Linguistics

Proceedings of the Conference and Workshops

COLING

Volume 29 (2022), No. 15

Proceedings of the 9th Workshop on Asian Translation (WAT2022)

The 29th International Conference on Computational Linguistics

October 17, 2022
Gyeongju, Republic of Korea

ISSN 2951-2093

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 language. 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 – WAT2021), WAT2022 will bring together machine translation researchers and users to try, evaluate, share and discuss brand-new ideas about machine translation. For the 9th WAT, we included several new translation tasks including Structured Document Translation Task, Video Guided Ambiguous Subtitling Task, Khmer Speech Translation Task, Two new translation tasks to the Restricted Translation task, Parallel Corpus Filtering Task, Bengali Visual Genome Task, 5 new languages to the Multilingual Indic Machine Translation Task and 1 new language to the Wikinews and Software Documentation Translation Task. We had 8 teams participate in the shared tasks. About 300 translation results were submitted to the automatic evaluation server, and selected submissions were manually evaluated. In addition to the shared tasks, WAT2022 also features research papers on topics related to machine translation, especially for Asian languages. The program committee accepted 4 research papers.

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 COLING2022 organizers for their help with administrative matters.

WAT 2022 Organizers

Organizing Committee:

Toshiaki Nakazawa, The University of Tokyo, Japan
Isao Goto, Japan Broadcasting Corporation (NHK), Japan
Hideya Mino, Japan Broadcasting Corporation (NHK), 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
Shohei Higashiyama, National Institute of Information and Communications Technology (NICT), Japan
Hiroshi Manabe, National Institute of Information and Communications Technology (NICT), Japan
Shantipriya Parida, Silo AI, Finland
Ondřej Bojar, Charles University, Czech Republic
Chenhui Chu, Kyoto University, Japan
Akiko Eriguchi, Microsoft, USA
Kaori Abe, Tohoku University, Japan
Yusuke Oda, LegalForce, Japan
Makoto Morishita, NTT, Japan
Katsuhito Sudoh, Nara Institute of Science and Technology (NAIST), Japan
Sadao Kurohashi, Kyoto University, Japan
Pushpak Bhattacharyya, Indian Institute of Technology Patna (IITP), India

Program Committee:

Chenhui Chu, Kyoto University, Japan
Sangjee Dondrub, Qinghai Normal University, China
Chao-Hong Liu, ADAPT Centre, Dublin City University, Ireland
Valentin Malykh, Huawei Noah's Ark Lab / Kazan Federal University, Russian Federation
Takashi Ninomiya, Ehime University, Japan
Katsuhito Sudoh, Nara Institute of Science and Technology (NAIST), Japan
Masao Utiyama, NICT, Japan
Xinyi Wang, Carnegie Mellon University, United States
Jiajun Zhang, Chinese Academy of Sciences, 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
Chen Zhang, National University of Singapore, Singapore

Invited talk: Machine translation of Turkic languages: Current approaches and Open challenges

Duygu Ataman

New York University

Abstract

Recent advances in neural machine translation have pushed the quality of machine translation systems to the point where they are becoming widely adopted to build competitive systems. However, there is still a large number of languages that are yet to reap the benefits of neural machine translation. In this context, we present a review of the neural machine translation technology and the results from a large-scale case study of the practical application of neural machine translation in the Turkic language family in order to realize the applicability of prominent architectures and learning methods, data sets as well as evaluation metrics in languages with different characteristics and under high-resource to extremely low-resource scenarios, in addition to identified limitations and promising directions for research to contribute to the extension of the applicability of translation technology in more languages and domains.

Table of Contents

Overview of the 9th Workshop on Asian Translation
Toshiaki Nakazawa, Hideya Mino, Isao Goto, Raj Dabre, Shohei Higashiyama, Shantipriya Parida, Anoop Kunchukuttan, Makoto Morishita, Ondřej Bojar, Chenhui Chu, Akiko Eriguchi, Kaori Abe, Yusuke Oda and Sadao Kurohashi 1

Comparing BERT-based Reward Functions for Deep Reinforcement Learning in Machine Translation
Yuki Nakatani, Tomoyuki Kajiwara and Takashi Ninomiya 37

Improving Jejueo-Korean Translation With Cross-Lingual Pretraining Using Japanese and Korean
Francis Zheng, Edison Marrese-Taylor and Yutaka Matsuo 44

TMU NMT System with Automatic Post-Editing by Multi-Source Levenshtein Transformer for the Restricted Translation Task of WAT 2022
Seiichiro Kondo and Mamoru Komachi 51

HwTscSU's Submissions on WAT 2022 Shared Task
Yilun Liu, Zhen Zhang, shimin tao, Junhui Li and Hao Yang 59

NICT's Submission to the WAT 2022 Structured Document Translation Task
Raj Dabre 64

Rakuten's Participation in WAT 2022: Parallel Dataset Filtering by Leveraging Vocabulary Heterogeneity
Alberto Poncelas, Johanes Effendi, Ohnmar Htun, Sunil Yadav, Dongzhe Wang and Saurabh Jain 68

NIT Rourkela Machine Translation(MT) System Submission to WAT 2022 for MultiIndicMT: An Indic Language Multilingual Shared Task
Sudhansu Bala Das, Atharv Biradar, Tapas Kumar Mishra and Bidyut Kumar Patra 73

Investigation of Multilingual Neural Machine Translation for Indian Languages
Sahinur Rahman Laskar, Riyanka Manna, Partha Pakray and Sivaji Bandyopadhyay 78

Does partial pretranslation can improve low ressourced-languages pairs?
raoul blin 82

Multimodal Neural Machine Translation with Search Engine Based Image Retrieval
ZhenHao Tang, XiaoBing Zhang, Zi Long and XiangHua Fu 89

Silo NLP's Participation at WAT2022
Shantipriya Parida, Subhadarshi Panda, Stig-Arne Grönroos, Mark Granroth-Wilding and Mika Koistinen 99

PICT@WAT 2022: Neural Machine Translation Systems for Indic Languages
Anupam Patil, Isha Joshi and Dipali Kadam 106

English to Bengali Multimodal Neural Machine Translation using Transliteration-based Phrase Pairs Augmentation
Sahinur Rahman Laskar, Pankaj Dadure, Riyanka Manna, Partha Pakray and Sivaji Bandyopadhyay 111

Investigation of English to Hindi Multimodal Neural Machine Translation using Transliteration-based Phrase Pairs Augmentation
Sahinur Rahman Laskar, Rahul Singh, Md Faizal Karim, Riyanka Manna, Partha Pakray and Sivaji Bandyopadhyay 117

Workshop Program

October 17, 2022 [UTC+9]

9:00–9:05 Welcome

Overview of the 9th Workshop on Asian Translation
Toshiaki Nakazawa, Hideya Mino, Isao Goto, Raj Dabre, Shohei Higashiyama, Shantipriya Parida, Anoop Kunchukuttan, Makoto Morishita, Ondřej Bojar, Chenhui Chu, Akiko Eriguchi, Kaori Abe, Yusuke Oda and Sadao Kurohashi

9:05–9:50 Invited Talk

Machine translation of Turkic languages: Current approaches and Open challenges
Duygu Ataman

9:50–10:30 Research Paper I

Comparing BERT-based Reward Functions for Deep Reinforcement Learning in Machine Translation
Yuki Nakatani, Tomoyuki Kajiwara and Takashi Ninomiya

Improving Jejueo-Korean Translation With Cross-Lingual Pretraining Using Japanese and Korean
Francis Zheng, Edison Marrese-Taylor and Yutaka Matsuo

10:30–11:00 Break

October 17, 2022 [UTC+9] (continued)

11:00–12:30 Shared Task I

Task Descriptions and Results: Restricted
Kaori Abe

TMU NMT System with Automatic Post-Editing by Multi-Source Levenshtein Transformer for the Restricted Translation Task of WAT 2022
Seiichiro Kondo and Mamoru Komachi

Task Descriptions and Results: Software
Raj Dabre

HwTscSU's Submissions on WAT 2022 Shared Task
Yilun Liu, Zhen Zhang, shimin tao, Junhui Li and Hao Yang

Task Descriptions and Results: SWSTR
Raj Dabre

NICT's Submission to the WAT 2022 Structured Document Translation Task
Raj Dabre

12:30–14:00 Lunch Break

14:00–15:20 Shared Task II

Task Descriptions and Results: Parallel Corpus Filtering
Makoto Morishita

Rakuten's Participation in WAT 2022: Parallel Dataset Filtering by Leveraging Vocabulary Heterogeneity
Alberto Poncelas, Johanes Effendi, Ohnmar Htun, Sunil Yadav, Dongzhe Wang and Saurabh Jain

Task Descriptions and Results: Indic
Shantipriya Parida

NIT Rourkela Machine Translation(MT) System Submission to WAT 2022 for MultiIndicMT: An Indic Language Multilingual Shared Task
Sudhansu Bala Das, Atharv Biradar, Tapas Kumar Mishra and Bidyut Kumar Patra

October 17, 2022 [UTC+9] (continued)

Investigation of Multilingual Neural Machine Translation for Indian Languages
Sahinur Rahman Laskar, Riyanka Manna, Partha Pakray and Sivaji Bandyopadhyay

15:20–16:00 Break

16:00–16:40 Research Paper II

Does partial pretranslation can improve low ressourced-languages pairs?
raoul blin

Multimodal Neural Machine Translation with Search Engine Based Image Retrieval
ZhenHao Tang, XiaoBing Zhang, Zi Long and XiangHua Fu

16:40–17:50 Shared Task III

Task Descriptions and Results: Multimodal
Shantipriya Parida

Silo NLP's Participation at WAT2022
Shantipriya Parida, Subhadarshi Panda, Stig-Arne Grönroos, Mark Granroth-Wilding and Mika Koistinen

PICT@WAT 2022: Neural Machine Translation Systems for Indic Languages
Anupam Patil, Isha Joshi and Dipali Kadam

English to Bengali Multimodal Neural Machine Translation using Transliteration-based Phrase Pairs Augmentation
Sahinur Rahman Laskar, Pankaj Dadure, Riyanka Manna, Partha Pakray and Sivaji Bandyopadhyay

Investigation of English to Hindi Multimodal Neural Machine Translation using Transliteration-based Phrase Pairs Augmentation
Sahinur Rahman Laskar, Rahul Singh, Md Faizal Karim, Riyanka Manna, Partha Pakray and Sivaji Bandyopadhyay

October 17, 2022 [UTC+9] (continued)

17:50–17:55 Closing