

IJCNLP 2017

The 4th Workshop on Asian Translation

Proceedings of the Workshop

November 27 - December 1, 2017
Taipei, Taiwan

©2017 Asian Federation of Natural Language Processing

ISBN 978-1-948087-06-3

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, WAT2015, and WAT2016), WAT2017 brings together machine translation researchers and users to try, evaluate, share and discuss brand-new ideas about machine translation. For the 4th WAT, we proudly include new domains: Newswire and Recipe in addition to scientific paper, patent, and mixed domain for the machine translation evaluation shared tasks. We had 12 teams who submitted their translation results, and about 300 submissions in total.

In addition to the shared tasks, WAT2017 also features scientific papers on topics related to the machine translation, especially for Asian languages. The program committee accepted 4 papers, which focus on neural machine translation, and construction and evaluation of language resources.

We are grateful to "SunFlare Co., Ltd." 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 IJCNLP 2017 organizers for their help with administrative matters.

WAT2017 Organizers

Organizers:

Toshiaki Nakazawa, Japan Science and Technology Agency (JST), Japan
Hideya Mino, National Institute of Information and Communications Technology (NICT), Japan
Chenchen Ding, National Institute of Information and Communications Technology (NICT), Japan
Shohei Higashiyama, National Institute of Information and Communications Technology (NICT), Japan
Isao Goto, Japan Broadcasting Corporation (NHK), Japan
Graham Neubig, Nara Institute of Science and Technology (NAIST), Japan
Hideto Kazawa, Google, Japan
Yusuke Oda, Nara Institute of Science and Technology (NAIST), Japan
Jun Harashima, Cookpad Inc., Japan
Sadao Kurohashi, Kyoto University, Japan
Ir. Hammam Riza, Agency for the Assessment and Application of Technology (BPPT), Indonesia
Pushpak Bhattacharyya, Indian Institute of Technology Bombay (IIT), India

Program Committee:

Chenhui Chu, JST, Japan
Fabien Cromières, JST, Japan
Hideto Kazawa, Google, Japan
Anoop Kunchookuttan, IIT Bombay, India
Qun Liu, Dublin City University, Ireland
Yvette Graham, Dublin City University, Ireland
Yang Liu, Tsinghua University, China
Liling Tan, Universität des Saarlandes, Germany
Masao Utiyama, NICT, Japan
Jiajun Zhang, Chinese Academy of Sciences, China

Technical Collaborators:

Luis Fernando D'Haro, Institute for Infocomm Research, Singapore
Rafael E. Banchs, Institute for Infocomm Research, Singapore
Haizhou Li, Institute for Infocomm Research, Singapore

Invited talk: Turning NMT Research into Commercial Products

Adrià de Gispert

SDL

Abstract

Recently, neural machine translation has revolutionised the field of machine translation, and now results in many research tasks keep improving every year. The new neural models have greatly improved translation quality, but have very different sorts of errors than the traditional statistical machine translation technology. An important challenge is to incorporate this technology improvement into commercial products and ensure that machine translation users get the best value while still keeping the product features they rely on for their work. SDL provides machine translation technology in a variety of products and markets. Our customers have expectations related to decoding speed, support for dictionaries and tags, and other functionality, so they can successfully integrate MT in their workflows. When it comes to commercialising MT, ensuring that these expectations are met is as important as improvements in BLEU score. In this talk I will focus on these important practical aspects in the context of the current NMT developments.

Biography

Dr. Adrià de Gispert is a senior research scientist at SDL Research, as well as a senior research associate at the Engineering Department in the University of Cambridge, UK. He received his PhD on Statistical Machine Translation from Universitat Politècnica de Catalunya (UPC, Barcelona) in 2007. Then he moved to Cambridge, where he has continued working in this field since, both in academia and in industry. He has published more than 30 major research papers on MT, and has contributed to the development of multiple state-of-the-art research and commercial machine translation engines, including phrase-based, syntax-based and neural. He is a Fellow of Clare College, Cambridge.

Table of Contents

<i>Overview of the 4th Workshop on Asian Translation</i> Toshiaki Nakazawa, Shohei Higashiyama, Chenchen Ding, Hideya Mino, Isao Goto, Hideto Kazawa, Yusuke Oda, Graham Neubig and Sadao Kurohashi	1
<i>Controlling Target Features in Neural Machine Translation via Prefix Constraints</i> Shunsuke Takeno, Masaaki Nagata and Kazuhide Yamamoto	55
<i>Improving Japanese-to-English Neural Machine Translation by Paraphrasing the Target Language</i> Yuuki Sekizawa, Tomoyuki Kajiwara and Mamoru Komachi	64
<i>Improving Low-Resource Neural Machine Translation with Filtered Pseudo-Parallel Corpus</i> Aizhan Imankulova, Takayuki Sato and Mamoru Komachi	70
<i>Japanese to English/Chinese/Korean Datasets for Translation Quality Estimation and Automatic Post-Editing</i> Atsushi Fujita and Eiichiro Sumita	79
<i>NTT Neural Machine Translation Systems at WAT 2017</i> Makoto Morishita, Jun Suzuki and Masaaki Nagata	89
<i>XMU Neural Machine Translation Systems for WAT 2017</i> Boli Wang, Zhixing Tan, Jinming Hu, Yidong Chen and xiaodong shi	95
<i>A Bag of Useful Tricks for Practical Neural Machine Translation: Embedding Layer Initialization and Large Batch Size</i> Masato Neishi, Jin Sakuma, Satoshi Tohda, Shonosuke Ishiwatari, Naoki Yoshinaga and Masashi Toyoda	99
<i>Patent NMT integrated with Large Vocabulary Phrase Translation by SMT at WAT 2017</i> Zi Long, Ryuichiro Kimura, Takehito Utsuro, Tomoharu Mitsuhashi and Mikio Yamamoto ...	110
<i>SMT reranked NMT</i> Terumasa Ehara	119
<i>Ensemble and Reranking: Using Multiple Models in the NICT-2 Neural Machine Translation System at WAT2017</i> Kenji Imamura and Eiichiro Sumita	127
<i>A Simple and Strong Baseline: NAIST-NICT Neural Machine Translation System for WAT2017 English-Japanese Translation Task</i> Yusuke Oda, Katsuhito Sudoh, Satoshi Nakamura, Masao Utiyama and Eiichiro Sumita	135
<i>Comparison of SMT and NMT trained with large Patent Corpora: Japio at WAT2017</i> Satoshi Kinoshita, Tadaaki Oshio and Tomoharu Mitsuhashi	140
<i>Kyoto University Participation to WAT 2017</i> Fabien Cromieres, Raj Dabre, Toshiaki Nakazawa and Sadao Kurohashi	146
<i>CUNI NMT System for WAT 2017 Translation Tasks</i> Tom Kocmi, Dušan Variš and Ondřej Bojar	154

<i>Tokyo Metropolitan University Neural Machine Translation System for WAT 2017</i> Yukio Matsumura and Mamoru Komachi	160
<i>Comparing Recurrent and Convolutional Architectures for English-Hindi Neural Machine Translation</i> Sandhya Singh, Ritesh Panjwani, Anoop Kunchukuttan and Pushpak Bhattacharyya	167

Workshop Program

November 27, 2017

9:00–9:20 **Welcome & Overview**

Overview of the 4th Workshop on Asian Translation

Toshiaki Nakazawa, Shohei Higashiyama, Chenchen Ding, Hideya Mino, Isao Goto, Hideto Kazawa, Yusuke Oda, Graham Neubig and Sadao Kurohashi

9:20–10:40 **Research Paper**

Controlling Target Features in Neural Machine Translation via Prefix Constraints

Shunsuke Takeno, Masaaki Nagata and Kazuhide Yamamoto

Improving Japanese-to-English Neural Machine Translation by Paraphrasing the Target Language

Yuuki Sekizawa, Tomoyuki Kajiwara and Mamoru Komachi

Improving Low-Resource Neural Machine Translation with Filtered Pseudo-Parallel Corpus

Aizhan Imankulova, Takayuki Sato and Mamoru Komachi

Japanese to English/Chinese/Korean Datasets for Translation Quality Estimation and Automatic Post-Editing

Atsushi Fujita and Eiichiro Sumita

10:40–11:00 **Coffee break**

11:00–12:00 **System description**

NTT Neural Machine Translation Systems at WAT 2017

Makoto Morishita, Jun Suzuki and Masaaki Nagata

XMU Neural Machine Translation Systems for WAT 2017

Boli Wang, Zhixing Tan, Jinming Hu, Yidong Chen and xiaodong shi

A Bag of Useful Tricks for Practical Neural Machine Translation: Embedding Layer Initialization and Large Batch Size

Masato Neishi, Jin Sakuma, Satoshi Tohda, Shonosuke Ishiwatari, Naoki Yoshinaga and Masashi Toyoda

November 27, 2017 (continued)

14:00–14:45 Invited talk

Turning NMT Research into Commercial Products

Adrià de Gispert

14:45–15:05 Poster booster (9 systems)

15:05–15:10 Commemorative photo

15:10–15:30 Coffee break

15:30–16:15 Poster presentation I (System description)

NTT Neural Machine Translation Systems at WAT 2017

Makoto Morishita, Jun Suzuki and Masaaki Nagata

Patent NMT integrated with Large Vocabulary Phrase Translation by SMT at WAT 2017

Zi Long, Ryuichiro Kimura, Takehito Utsuro, Tomoharu Mitsuhashi and Mikio Yamamoto

SMT reranked NMT

Terumasa Ehara

Ensemble and Reranking: Using Multiple Models in the NICT-2 Neural Machine Translation System at WAT2017

Kenji Imamura and Eiichiro Sumita

A Simple and Strong Baseline: NAIST-NICT Neural Machine Translation System for WAT2017 English-Japanese Translation Task

Yusuke Oda, Katsuhito Sudoh, Satoshi Nakamura, Masao Utiyama and Eiichiro Sumita

Comparison of SMT and NMT trained with large Patent Corpora: Japio at WAT2017

Satoshi Kinoshita, Tadaaki Oshio and Tomoharu Mitsuhashi

November 27, 2017 (continued)

16:15–17:00 Poster presentation II (System description)

A Bag of Useful Tricks for Practical Neural Machine Translation: Embedding Layer Initialization and Large Batch Size

Masato Neishi, Jin Sakuma, Satoshi Tohda, Shonosuke Ishiwatari, Naoki Yoshinaga and Masashi Toyoda

Kyoto University Participation to WAT 2017

Fabien Cromieres, Raj Dabre, Toshiaki Nakazawa and Sadao Kurohashi

CUNI NMT System for WAT 2017 Translation Tasks

Tom Kocmi, Dušan Variš and Ondřej Bojar

XMU Neural Machine Translation Systems for WAT 2017

Boli Wang, Zhixing Tan, Jinming Hu, Yidong Chen and xiaodong shi

Tokyo Metropolitan University Neural Machine Translation System for WAT 2017

Yukio Matsumura and Mamoru Komachi

Comparing Recurrent and Convolutional Architectures for English-Hindi Neural Machine Translation

Sandhya Singh, Ritesh Panjwani, Anoop Kunchukuttan and Pushpak Bhattacharyya

17:00– Closing

