ICON 2020

17th International Conference on Natural Language Processing

Proceedings of the Adap-MT 2020 Shared Task

December 18 - 21, 2020
Indian Institute of Technology Patna, India

Introduction

These shared task proceedings concluded the shared task on Low Resource Domain Adaptation for Indic Machine Translation, named as ADAP-MT 2020, launched on 7th October 2020. The shared task was collocated with the 17th International Conference on Natural Language Processing (ICON 2020), held at IIT-Patna, India. The goal of the shared task was to show how MT systems trained on general domains perform on Indic Languages and low resource domain adaptation using a limited domain-specific parallel corpus.

Two subtasks were part of this shared task. In the first subtask, the participants were asked to develop MT systems for the General domain. The second subtask required the development of MT systems for specified domains - AI, Chemistry utilizing general domain parallel data and very limited domain-specific data for domain adaptation. Parallel corpora of three language pairs - English - Hindi, English - Telugu and Hindi - Telugu were released for the shared task.

We received five system submissions and system description papers. Each system description paper was reviewed by two members of the reviewing committee – all papers were accepted. The submitted systems were evaluated using BLEU scores.

Statistical MT and Neural MT were the two kinds of models used by the participants. Subword level NMT models with Byte Pair Encoding with self - Attention were mostly used. Participants also augmented the training data with techniques like oversampling of domain-specific data and mixed fine-tuning. We would like to thank the ICON-2020 organizers, the shared task participants, the authors, and the reviewers for making this shared task successful.

Shared task page: `http://ssmt.iiit.ac.in/machinetranslation`
Main conference page: `https://www.iitp.ac.in/~ai-nlp-ml/icon2020/index.html`

Organizing Committee:

Dipti Misra Sharma (IIIT-Hyderabad)
Asif Ekbal (IIT-Patna)
Karunesh Arora (C-DAC, Noida)
Sudip Kumar Naskar (Jadavpur University)
Dipankar Ganguly (C-DAC, Noida)
Sobha L (AUKBC-Chennai)
Radhika Mamidi (IIIT-Hyderabad)
Sunita Arora (C-DAC, Noida)
Pruthwik Mishra (IIIT-Hyderabad)
Vandan Mujadia (IIIT-Hyderabad)

Table of Contents

Shared Task Program

Monday, December 21, 2020

+ 14:00 - 14:30 **Talk by Sobha L, AUKBC-Chennai**

+ 14:30 - 14:45 Shared Task Overview

Presentations

14:50 - 15:00   *Terminology-Aware Sentence Mining for NMT Domain Adaptation: ADAPT's Submission to the Adap-MT 2020 English-to-Hindi AI Translation Shared Task*
Rejwanul Haque, Yasmin Moslem and Andy Way

15:03 - 15:13   *AdapNMT : Neural Machine Translation with Technical Domain Adaptation for Indic Languages*
Hema Ala, Dipti Misra Sharma

15:16 - 15:26   *Domain Adaptation of NMT models for English-Hindi Machine Translation Task at AdapMT ICON 2020*
Ramchandra Joshi, Rushabh Karnavat, Kaustubh Jirapure, Raviraj Joshi

15:29 - 15:39   *JUNLP@ICON2020: Low Resourced Machine Translation for Indic Languages*
Sainik Kumar Mahata, Dipankar Das, Sivaji Bandyopadhyay

15:42 - 15:52   *MUCS@Adap-MT 2020: Low Resource Domain Adaptation for Indic Machine Translation*
Asha Hegde, H. L. Shashirekha