Overview of WAT2014

Participants List

- WASUIPS
- NICT
- Sense
- SAS_MT
- BJTUNLP
- TMU
- WEBLIO-EJ1
- Kyoto-U
- EIWA
- Saarland University & SAS Research and Development Co., Ltd
- National Institute of Information and Communication Technology
- TOSHIBA
- WASUIPS

Manual (HUMAN) and Automatic (BLEU) Evaluation Results and Their Correlations

HUMAN score

- Pairwise evaluation compared to the baseline using crowdsourcing
- Reduce the cost, time, number of sentences to be evaluated
- Enable the evaluation of new translation results after the workshop
- Reference translations are not shown

Inter Annotator Agreement

- 1,000 sentences are randomly selected (paragraph-based) from the test set
- 400 sentences are randomly selected by the voting of 3 judgments
- Each pair is evaluated 3 different workers
- To guarantee the quality of the evaluation, each pair is evaluated 3 different workers
- The evaluation result of each pair is decided pairwise evaluation compared to the baseline using crowdsourcing

Case Study

- Submission ID
- BLEU
- RIBES
- HUMAN
- Forest Input

Thank you very much for attending WAT2014
Stay tuned for the next WAT workshop!