ISCA Archive
http://www.isca-speech.org/archive

INTERSPEECH 2008
9th Annual Conference of the International Speech Communication Association

interspeech 2008 incorporating SST 08

Brisbane, Australia
September 22-26, 2008

Data Selection and Smoothing in an Open-Source System for the 2008 NIST Machine Translation Evaluation

Holger Schwenk, Yannick Esteve

LIUM, France

This paper gives a detailed description of a statistical machine translation system developed for the 2008 NIST open MT evaluation. The system is based on the open source toolkit Moses with extensions for language model rescoring in a second pass. Significant improvements were obtained with data selection methods for the language and translation model. An improvement of more than 1 point BLEU on the test set was achieved by a continuous space language model which performs the probability estimation with a neural network. The described system has achieved a very good ranking in the 2008 NIST open MT evaluation.

Full Paper

Bibliographic reference. Schwenk, Holger / Esteve, Yannick (2008): "Data selection and smoothing in an open-source system for the 2008 NIST machine translation evaluation", In *INTERSPEECH-2008*, 2727-2730.