

Tutorial Submission

Statistical Machine Translation with Open Source Software, Hieu Hoang and Nadir Durrani

This tutorial will introduce the open source tools such as the Moses toolkit and stress its functionality. It is a hands-on tutorial more geared towards using the tool, rather than a descriptions of the details of the methods. The tutorial will cover the following topics: overview of the training and tuning pipeline; working with an experimental management system; analyzing the output; word alignment toolkits; input modalities: words, confusion networks, lattices; output modalities: 1-best, n-best, search graph; translation models: phrasal, hierarchical, syntactified; language models: binarization, quantization, randomization; decision rules: MAP, MBR, lattice MBR, consensus; trade-offs between speed and quality; trade-offs between speed and memory efficiency; incremental updating of the translation model; adding features to the decoder.

Hieu Hoang, University of Edinburgh

Hieu Hoang is a postgraduate researcher at the University of Edinburgh working on improving machine translation with the addition of linguistic information. He is a founding member of the team that created the Moses toolkit and continues to be a major contributor and maintainer of the system. He has degrees in Computer Science and Machine Learning (London) and has over 10 years experience as a software developer. He holds a PhD from the University of Edinburgh

Matthias Huck, University of Edinburgh

Matthias Huck is a research associate at the University of Edinburgh. He is working on statistical machine translation since 2008 and has published more than 30 journal, conference, or workshop papers. His research interests cover a wide range of topics in machine translation, including phrase-based, hierarchical, and syntax-based translation, as well as system combination. Matthias is an active Moses developer.

Prior to joining the Statistical Machine Translation group at the University of Edinburgh in 2013, Matthias was a member of the Human Language Technology and Pattern Recognition group at RWTH Aachen University, where he was one of the core developers of Jane, RWTH's open source statistical machine translation