

Title: Measures of Machine Translation Quality

Author: Matouš Macháček

Department: Institute of Formal and Applied Linguistics

Supervisor: RNDr. Ondřej Bojar, Ph.D.

Abstract: We explore both manual and automatic methods of machine translation evaluation. We propose a manual evaluation method in which annotators rank only translations of short segments instead of whole sentences. This results in easier and more efficient annotation. We have conducted an annotation experiment and evaluated a set of MT systems using this method. The obtained results are very close to the official WMT14 evaluation results. We also use the collected database of annotations to automatically evaluate new, unseen systems and to tune parameters of a statistical machine translation system. The evaluation of unseen systems, however, does not work and we analyze the reasons. To explore the automatic methods, we organized Metrics Shared Task held during the Workshop of Statistical Machine Translation in years 2013 and 2014. We report the results of the last shared task, discuss various metaevaluation methods and analyze some of the participating metrics.

Keywords: machine translation, evaluation, automatic metrics, annotation