

Improving Rule-Based Machine Translation by Using Statistical Syntactic Rules

Hakimeh Fadaei

Heshaam Faili

Farnaz Ghasemi

School of Electrical and Computer Engineering, University of Tehran, Tehran, Iran

ABSTRACT

Rule-based machine translation uses a set of linguistic rules in the process of translation. The results of these systems are usually better than the results of statistical models from grammatical and word order perspective, But it has been shown that statistical models are more powerful in selecting proper words and generating more fluent translations. In this paper our goal is to improve the word choice in rule-based machine translation. This is done by a set of lexical syntactic rules based on Tree Adjoining Grammar. These probabilistic rules are statistically extracted from a large parallel corpus. In the proposed system, the input sentence is first reordered by a rule-based system, and then the decoding is carried out monotonically by using dynamic programming. In this system the best translation is chosen based on the extracted rules and the language model score. The experiments on English-Persian translation showed that the proposed method resulted in an improvement of 1.3 in BLEU score in comparison to our baseline rule-based method.

Keywords: Hybrid Machine Translation, Rule-Based Machine Translation, Statistical Rules, Lexical Syntactic Rules, Tree Adjoining Grammar.