Fuzzy Matching for Tree-based Machine Translation

Xiong Hao  Liu Yang  Liu Qun
Institute of Computing Technology, Chinese Academy of Sciences, Beijing 100190
E-mail: {xionghao, yliu, liuqun} @ ict.ac.cn

Abstract: Previous related work of tree-based models treat rules as strings and then match rules using string matching algorithm. However, the performance of tree-based models is largely depended on the parsing results, and for some languages, the precision of current parser is still far from state-of-the-art. So two rules with one different tag causing by parsing errors seems to be unmatchable. Under exact matching strategy, the size of available rules is implicitly scarce especially in tree-to-tree models, in which the performance is still unacceptable. In this paper, we present a tree kernel based fuzzy matching algorithm which computes the similarity between different rules. Experimental results on NIST 2005 Chinese-to-English test set show that our system achieve an absolute improvement of 1.3% in term of BLEU score over string matching system. Furthermore, when using the packed forest, our method still gets a relative improvement of 0.7 BLEU score.

Keywords: tree kernel, tree-to-string model, statistical machine translation, fuzzy matching