

Abstract

Guillem Gascó Mora and Joan Andreu Sánchez Peiró. Syntax Augmented Inversion Transduction Grammars for Machine Translation

In this paper we propose a novel method for inferring an Inversion Transduction Grammar (ITG) with source (or target) language linguistic information from a bilingual parallel corpus. Our method combines bilingual ITG parse trees with monolingual linguistic source (or target) trees in order to obtain a Syntax Augmented ITG (SAITG). The use of a modified parsing algorithm for bilingual parsing with bracketing information makes possible that each bilingual subtree have a correspondent subtree in the monolingual parsing. In addition, several binarization techniques have been tested for the resulting SAITG. In order to evaluate the effects of the use of SAITGs in Machine Translation tasks, we have used them in an ITG-based machine translation decoder. The decoder is a hybrid machine translation system that combines phrase-based models together with syntax-based translation models. The formalism that underlies the whole decoding process is a Chomsky Normal Form Stochastic Inversion Transduction Grammar (SITG) with phrasal productions and a log-linear combination of probability models. The decoder uses a CYK-like algorithm that combines the translated phrases inversely or directly in order to get a complete translation of the input sentence. The results obtained using SAITGs with the decoder for the IWSLT-08 Chinese-English machine translation task produce significant improvements in BLEU and TER.