

Statistical Machine Translation Based on Predicate-Argument Structure

Chengqing Zong

Institute of Automantion, Chinese Academy of Sciences
No. 95, Zhong Guan Cun East Road, Beijing 100190, China
cqzong@lpr.ia.ac.cn

Abstract

As we have well known that it is always a basic requirement for statistical machine translation (SMT) to maintain semantic equivalence between a source sentence and its translation. However, nearly all of the existing translation models do not deal with the semantic structure between two languages at all. In this talk, I will present a novel translation method based on semantically-motivated framework, using predicate-argument structure (PAS). Generally, PAS depicts the semantic relation between a predicate and its associated arguments, and it always indicates the semantic frame and skeleton structure of a sentence. Thus, we believe the PAS would be much beneficial for machine translation in grasping the semantics of sentences. Furthermore, after analysis of the weakness of PAS representation during translation, I will propose a concept of syntax-complemented PAS (SC-PAS). It effectively overcomes the drawback of the prevalent gaps in PAS and provides more useful knowledge for SMT.

We also call the semantically-motivated framework as Analysis-Transformation-Translation (ATT) framework, which is just based on the PAS and SC-PAS. As the following figure shows, this framework divides the whole translation process into three steps: (1) Analysis: to analyze the source sentences and obtain their PASs (or SC-PASs) automatically; (2) Transformation: to convert the source-side PASs (or SC-PASs) to target side by predicate-aware transformation rules; (3) Translation: this step is further divided into two parts: (a)element translation is to translate each element of PAS (or SC-PAS); (b)translation by global reordering is to combine the resulting translation candidates to translate the entire structure. By taking advantage of PAS (or SC-PAS), the ATT framework can well keep the semantic structure consistency of the source language and the target language and consequently show the great potential to improve translation quality.

