Improving Phrase-based Machine Translation using Splitting Clause and Phrase Reordering

Name: Nguyen Vinh Van
Supervisor: Prof. Akira Shimazu

March 4 2008

1 Aim of Research

Phrase-based Statistical Machine Translation (PSMT) systems represent recently the state-of-the-art in statistical machine translation. However, these phrase-based models have some limitations. Firstly, with these models PSMT usually are powerful in word reordering within short distance, however, long distance reordering is still problematic. Secondly, syntactic transformations in the source or target languages are not captured. Consequently, our research aim focuses on exploiting and supplying linguistic knowledge to a PSMT system.

2 Proposed approach

Firstly, we consider the clause splitting in more detail. We find the very long and complicated sentences which are hard and costly to translate. Splitting these sentences into a set of smaller clauses could report many benefits for translation.

Secondly, reordering problem (global reordering) is one of the major problems in machine translation, since different languages have different word order requirements. We focuses on researching the ordering problem and aiming to improve both the quality of translation and computation time for decoding. Our approach is a global reordering model.

3 Progress of 2007

For the first problem, we present the CRFs-based framework model for Clause splitting. We use rich linguistic knowledge and a new bottom-up dynamic algorithm for decoding. The experiments show that our results are competitive as the previous results. The result is presented in the paper[1][2].

For the second problem, we present the new method for reordering in phrase based statistical machine translation. The experimental results with English-Vietnamese pair show that our method outperforms better both the accuracy and speed than the baseline PSMT.

4 Future Direction

We will to investigate to modifications of appropriate learning algorithms into the first and second problem. The implementations and experiments will apply for English-Japanese and English-French. Another work, we will integrate clause splitting into the machine translation system.

5 Publication

Journal paper

[1] Nguyen, V.V, Nguyen, L.M, A. Shimazu. "Clause Splitting with Conditional Random Fields", to be submitted.

Conference paper

[2] Nguyen, V.V, Nguyen, L.M, A. Shimazu. "Using Conditional Random Fields for Clause Splitting", In Proceedings of Pacling-07, pp. 58-65.