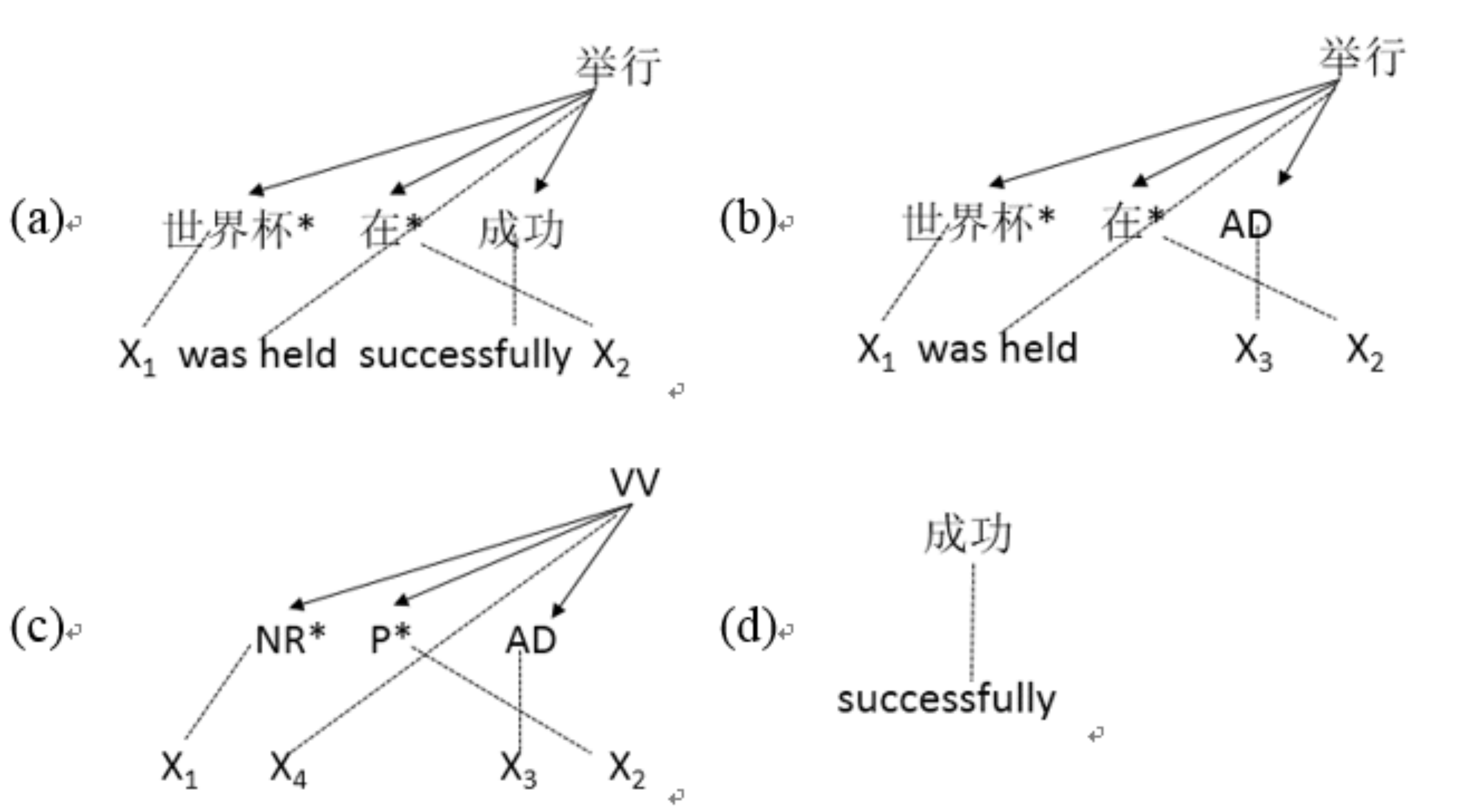
A Dependency-to-String Model for Chinese-Japanese SMT System

Hua Shan Lu Bai Te Luo Yujie Zhang
School of Computer and Information Technology, Beijing Jiaotong University
No.3 Shangyuancun, Haidian District, Beijing, China, 100044
{13120422, 13120397, 14120472 , yjzhang} @bjtu.edu.cn

Introduction

- Our system employs a dependency-to-string translation method for Chinese-Japanese SMT.
  - ➢ Motivated by that dependency grammar holds both syntactic and semantic knowledge.
- Our system achieves a BLEU of 34.87 and a RIBES of 79.25 on the Chinese-Japanese translation task in the official evaluation.
  - ➢ Our dependency-to-string model improved the BLEU score by 0.62 and the RIBES score by 0.31 on the test set.
  - ➢ The evaluation results illustrated that the translation system based on the dependency-to-string model is effective.

Dependency-to-String Grammar



The translation rules of the dependency-to-string model also act as reordering rules, which are classified into two types:

-HDR rules, the source side is generalized HDR fragments and the target sides is strings.

eg. $(x_1$:NR$)(x_2$:P$)(x_3$:AD$)$举行 → x_1 held x_3 x_2

-H rules, the source side is a word and the target side is words or strings.

eg. $(x_1$:世界杯$)(x_2$:在$)(x_3$:AD$)$举行 → x_1 held x_3 x_2

Rule Acquisition

We implemented the rule acquisition as follows:

1) Tree annotation: annotate the necessary information on each node of dependency trees for translation rule acquisition.
2) Identification of acceptable HDR fragments: identify HDR fragments from the annotated trees for HDR rules generation.
3) HDR rules generation: generate a set of HDR rules according to the identified acceptable HDR fragments

Decoding

Our decoder is based on bottom up chart parsing algorithm that convert the input dependency structure into a target string. It finds the best derivation among all possible derivations D. Given a source dependency structure T, the decoder traverses each internal node n of T in post-order. And we process it as follows.

1) If n is leaf node, it checks the translation rules H and uses the matched rules to generate candidate translation.
2) If n is a internal node, it enumerates all instances of the clauses or phrases of the HDR fragment rooted at n, and checks the translation rules for matched translation rules. If there is no matched rules, wo construct a pseudo translation rule according to the word order of the HDR fragment in the source side.
3) Make use of Cube Pruning algorithm to generate the candidate translation for the node n.

Translation Model

For a given source language dependency tree T, the model may generate more than one derivations D that convert a source dependency tree T into a target string e, thus producing a large amount of candidate translations. We adopt a general log-linear model to evaluate the candidate translations as:

$$P(D) \propto \prod \varphi_i(D)^{\lambda_i}$$

where φ_i are feature funcitons defined on derivation D and λ_i are the feature weights.

Our system used seven features as follows:

1) translation probabilities: P(t|s) and P(s|t)
2) lexical translation probabilities: $P_{lex}(t|s)$ and $P_{lex}(s|t)$
3) rule penalty: exp(-1)
4) target word penalty: exp(|e|)
5) language model: $P_{lm}(s)$

Experiment and Evaluation

We used Stanford Word Segmenter for Chinese word segmentation, with the standard of CTB, and used Stanford Parser for Chinese dependency parsing. We used JUMAN for Japanese word segmentation and used SRI Language Modeling Toolkit for training a 4-gram language model on the Japanese corpus preprocessed. We obtained the word alignments by running GIZA++ on the corpus in both directions and applying "grow-diag-and" refinement. We make use of MERT to tune the feature weights in order to maximize the system's BLEU score on the development set.

System	Rule #	BLEU	RIBES
Baseline	35M	34.25	78.94
Ours	8.8M	34.87	79.25

The Table shows the number of the extracted translation rules and the translation performance on the test data. Furthermore, we implemented a MOSES PBSMT system as the baseline for comparison. Ours performed better than Baseline by using only a small size of translation rules that is about one fourth of that of Baseline.