Using Syntax to Improve Word Alignment Precision for Syntax-Based Machine Translation

Author: Victoria Fossum et al.

Presenter: Jae Dong Kim

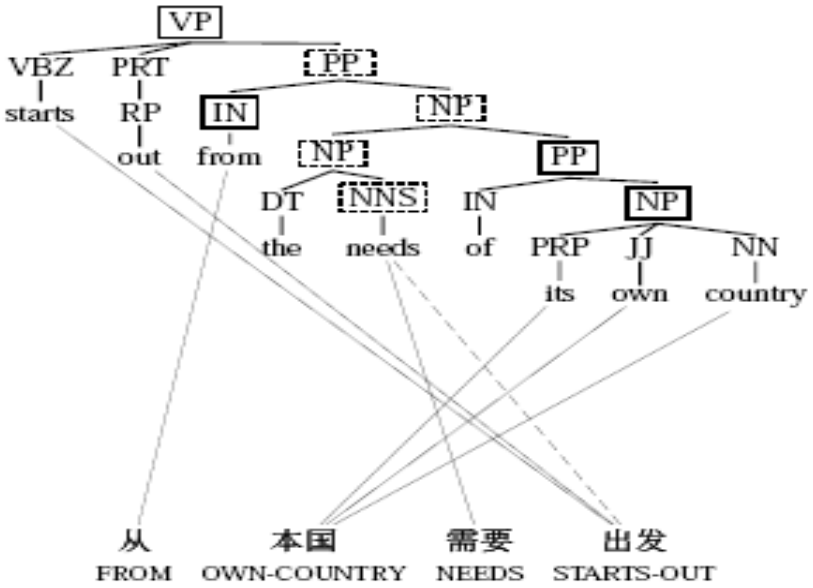
March 17, 2010

Introduction

- GIZA++ union
  - state-of-the-art syntax-based statistical MT (Galley et al., 2006)
  - hierarchical phrase-based system(Hiero)
- GIZA++ refined
  - phrase-based SMT (Och, Koehn)
- GIZA++ alignment improvement
  - union → deleting links
  - intersection → adding links

Introduction

- GIZA++ union
  - low precision
  - incorrect links
    - large size rules
    - few rules
    - rules with poor generalization ability

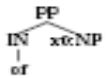


Rules Extracted Using GIZA++ Union Alignments

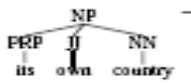
R1: VP → x0 x1 需要 出发

R2: IN → 从

R3: PP → x0



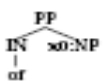
R4: NP → 本国



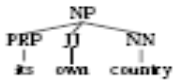
Rules Extracted After Deleting Dotted Link

R2: IN → 从

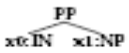
R3: PP → x0



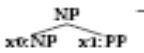
R4: NP → 本国



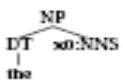
R5: PP → x0 x1



R6: NP → x1 x0



R7: NP → x0



R8: NNS → 需要

R9: VP → x0 出发

This work

- GIZA++ union → link deletion
  - supervised
  - features
    - size of largest rule
    - total number of rules
    - structural features
    - lexical features

Link Deletion Algorithm

- Definition
  - ***branch***: a contiguous 1-to-many alignment
  - A and A' are ***neighbor***s if they differ only by the deletion a link or branch
- For each sentence
  - $A = A_{initial}$ (GIZA++ union)
  - Score(A) = $\sum_{i=0}^{n} \lambda_i \cdot h_i(A, parse(e), f)$
  - Considering all A', greedily deleting the link *l* or branch of links *b* maximizing the score of the resulting alignment A' = A*l* or A'=A*b*
  - Do until no further increase of score is possible

Features

- Syntactic
  - ruleCount
  - sizeOfLargestRule
    - # internal nodes

Features

- Structural
  - wordUnaligned
    - # unaligned words
  - one-to-many Links
    - # 1-to-many in GIZA++ union > # 1-to-many in gold standard
  - zeroNeighbors
  - oneNeighbor
  - twoNeighbors
-

Features

- Lexical
  - highestLexProbRank
    - # max-probable links
- History
  - linksDeleted
    - a link or a branch is delete in each step
  - stepsTaken

Constraints

- Protecting Refined Links from Deletion
  - to keep higher precision
- Stoplist
  - Alignment links for the most common words that have no real translational equivalent are removed in pre-processing
  - all links in the set {a, an, the} X {de, le} were deleted from $A_{initial}$

Perceptron Training

- Averaged perceptron learning
  - weight for the most informative feature (ruleCount) ← 1.0, the others ← 0.0
  - In each pass over the training set
    - learn weights for each sentence

$$\lambda_i = \lambda_i + h_i^{A_{oracle}} - h_i^{A_{1-best}}$$

    - average the weights

Experimental Setup

- Language Pairs
  - two Chinese-English data sets
  - one Arabic-English data set
- Gold standard
  - each set contains a small subset of manually aligned sentence pairs → split into a training set and a test set
- Evaluation
  - alignment quality
  - translation quality

Evaluation Metrics

- Weighted Fully-Connected F-Measure
  - AER is not directly related to BLEU
  - F-measure(H^+)=(α /precision(H^+) + (1- α)/recall(H^+))$^{-1}$
  - α =0.5 for Cn/En and α =0.1 for Ar/En
- Rule F-Measure
  - P,R,F against rules extracted from gold alignment and gold parses
- BLEU

Data

Language	Train	Test
Chinese-English *A*	400	400
Chinese-English *B*	1500	1500
Arabic-English	1500	1500

Language	Train	Tune	Test1	Test2
Chinese-English *A*	9.8M/newswire	25.9k/NIST02	29.0k/NIST03	–
Chinese-English *B*	12.3M/newswire	42.9k/newswire	42.1k/newswire	–
Arabic-English	174.8M/newswire	35.8k/NIST04-05	40.3k/NIST04-05	53.0k/newswire

Results

Language	Alignment	Prec	Rec	α	F-measure	Links Del/ Sent	Grammar Size	BLEU Tune	BLEU Test1	BLEU Test2
Chi-Eng A	GIZA++ union	54.76	**75.38**	0.5	63.44	–	23.4M	41.80	41.17	–
Chi-Eng A	GIZA++ union + link deletion	**79.59**	71.16	0.5	**75.14**	4.77	59.7M	**43.06**	**41.93**	–
Chi-Eng B	GIZA++ union	36.61	**66.28**	0.5	47.16	–	28.9M	39.59	41.39	–
Chi-Eng B	GIZA++ union + link deletion	**65.52**	59.28	0.5	**62.24**	4.19	73.0M	**40.97**	**41.88**	–
Ara-Eng	GIZA++ union	35.34	**84.05**	0.1	73.87	–	52.4M	54.73	50.9	38.16
Ara-Eng	GIZA++ union + link deletion	**52.68**	79.75	0.1	**75.85**	1.35	64.9M	**55.57**	**51.08**	**38.72**

Table 3: Results of link deletion. Weighted fully-connected alignment f-measure is computed on alignment test sets (Table 1); BLEU score is computed on translation test sets (Table 2).

Results

Alignment	Parse	Rule			
		Precision	Recall	F-measure	Total Non-Unique
gold	gold	100.00	100.00	100.00	12,809
giza++ union	collins	**50.49**	44.23	47.15	11,021
giza++ union+link deletion, α=0.5	collins	47.51	53.20	**50.20**	13,987
giza++ refined	collins	44.20	**54.06**	48.64	15,182

Table 4: Rule precision, recall, and f-measure of rules extracted from 400 sentence pairs of Chinese-English data

Results

- Link deletion
  - Translation quality
    - GIZA++ union & link deletion > GIZA++ union
  - Alignment quality
    - improves fully-connected alignment precision
    - decreases fully-connected alignment recall
    - increases weighted f-measure overall

Discussion

- Why is recall more important for Arabic-English?
  - Greg used the *refined* one in WMT10 for French-English
- Rule recall for GIZA++ refined is higher than this approach → how about rule filtering on rules extracted from GIZA++ refined?

Backup (1)

- Fully-connected links
  - to overcome the same translation with different F