

IJCNLP 2008
Hyderabad, India

Refinements in BTG-based Statistical Machine translation

Deyi Xiong, Min Zhang, Aiti Aw
Institute for Infocomm Research
Haitao Mi, Qun Liu and Shouxun Lin
Institute of Computing Technology

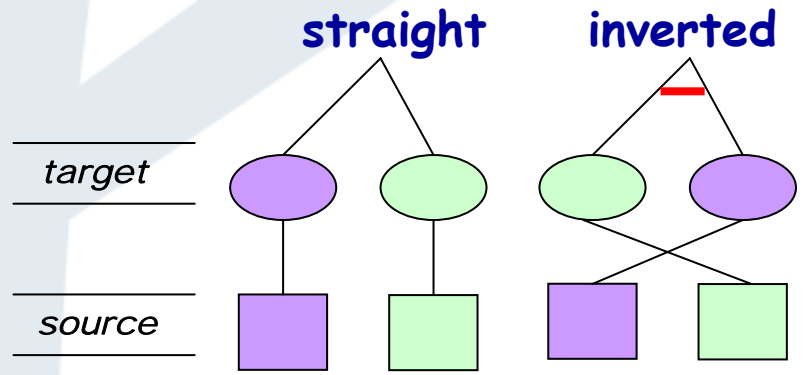
dyxiong@i2r.a-star.edu.sg

Outline

- Brief introduction of BTG-based SMT
- Major issues of BTG-based SMT
- Proposed refinements for BTG-based SMT
- Evaluation of proposed refinements
- Conclusions

BTG-based SMT: BTG rules

type		BTG rules
Merging rules	straight	$A \rightarrow [A^1, A^2]$
	inverted	$A \rightarrow \langle A^1, A^2 \rangle$
Lexical rule		$A \rightarrow x/y$



BTG-based SMT: Stochastic Model

- Model: TM, RM, LM, WM

$$e^* = \underset{e=Y_e(D)}{\operatorname{argmax}} P_T(r_{1..n_l}^l) \cdot P_R(r_{1..n_m}^m)^{\lambda_R} \cdot P_L(e)^{\lambda_L} \cdot \exp(|e|)^{\lambda_w}$$

- Translation model: P_T

$$P_T(r_{1..n_l}^l) = \prod_1^{n_l} P(r_i^l) = \prod_1^i P(x_i | y_i)^{\lambda_1} P(y_i | x_i)^{\lambda_2} P_{lex}(x_i | y_i)^{\lambda_3} P_{lex}(y_i | x_i)^{\lambda_4} \exp(1)^{\lambda_5}$$

- Reordering Model: P_R

$$P_R(r_{1..n_m}^m) = \prod_1^{n_m} P(r^m)$$

$$\text{flat} : P(r^m) = f(o, A^1, A^2) = \begin{cases} p_m & o = \text{straight} \\ 1 - p_m & o = \text{inverted} \end{cases}$$

$$\text{MaxEnt} : P(r^m) = f(o, A^1, A^2) = p_\theta(o | A^1, A^2) = \frac{\exp(\sum_i \theta_i h_i(o, A^1, A^2))}{\sum_o \exp(\sum_i \theta_i h_i(o, A^1, A^2))}$$

Major issues of BTG-based SMT

- Phrase reordering
 - Flat reordering: using prior probabilities without considering phrase content
 - MaxEnt reordering: boundary words predict wrong swappings
- Computational cost
 - Time complexity with m-gram language model used:
 $O(n^{3+4(m-1)})$

Refinements

- Reordering heuristics
 - Swapping window
 - Punctuation restriction
- Phrases with special tags
 - Indicating beginning and ending of sentence

- Prevent incorrect swapping
- Reduce search space

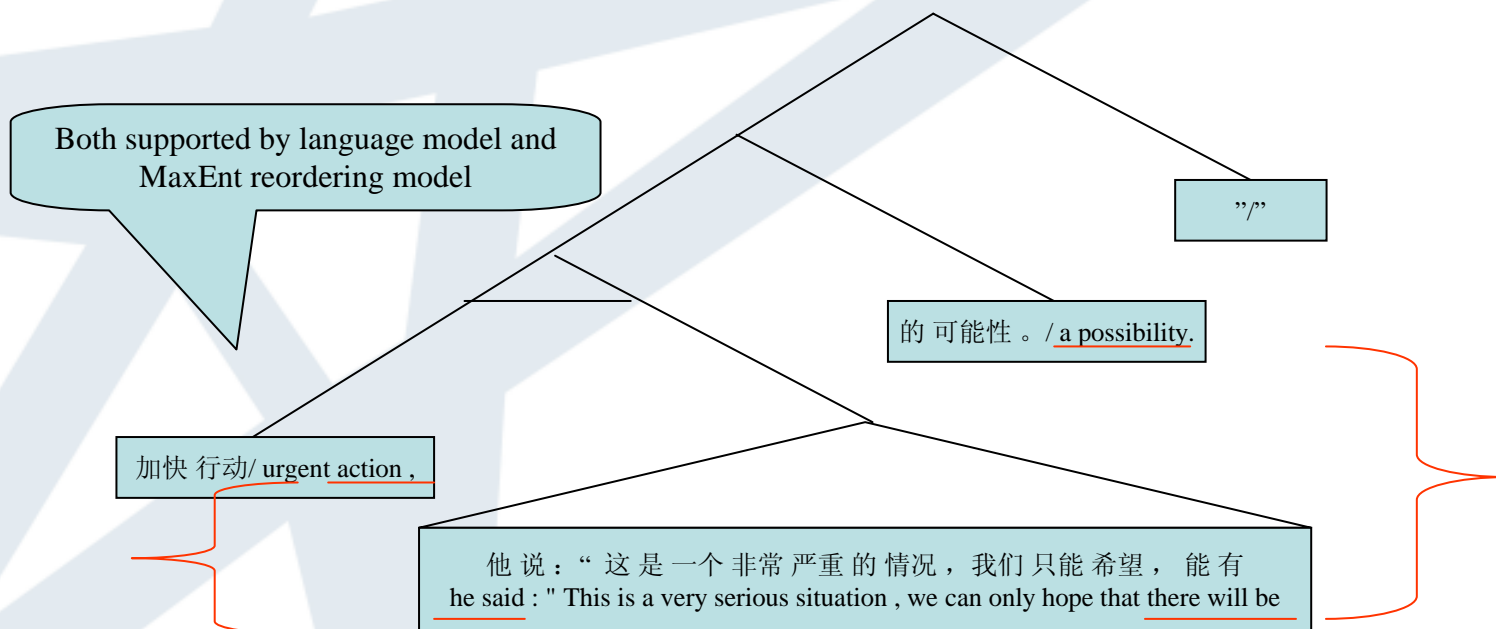
Why reordering heuristics:

Two findings related to reordering

- Incorrect swappings of two large neighboring blocks on the target side
- Incorrect swappings related to punctuation marks (PM)
 - Balanced PM (quotes, parentheses)
 - Source sequences within a pair of balanced PM is not kept within this PM after translation ✗
 - Structured PM (commas, semi-colons)
 - Source words are not translated monotonously relative to their adjacent PMs ✗

An example

- **Chinese:** 他说：“这是一个非常严重的情况，我们只能希望，能有加快行动的可能性。”
- **Ref:** He said: “This is a very serious situation. We can only hope that it is possible to speed up the operation.”
- **Bruin:** urgent action, he said: ``This is a very serious situation, we can only hope that there will be a possibility .”



How to address these problems?

- Two reordering heuristics:
 - Swapping window
 - Preventing long-distance reordering
 - Punctuation restriction
 - Using grammatical constraints of punctuation marks on word order around them

Swapping window

- The inverted rule is allowed in a swapping window
 - Activate $A \rightarrow \langle A^1, A^2 \rangle$ if $|A_s^1| + |A_s^2| <$ pre-defined swapping window size (m)
 - Any non-monotone combination beyond the swapping window size is prohibited
- Why?
 - Most reorderings are within a small window size according to our statistics on Chinese-to-English training corpus
- Another benefit: reducing search space
 - Monotone search span: $O((n-m)^2)$

Punctuation restriction

- If any punctuation marks, such as quotes, commas, parentheses, are included on the source side of two neighboring blocks,
 - They are merged with straight order
- Why?
 - Words around punctuation marks are always reordered monotonously.

Combined them together

- Swapping window + punctuation restriction:
 - Activate the inverted rule if: $|A_s^1| + |A_s^2| < m \ \&\& \ P \cap (A_s^1 \cup A_s^2) = \phi$

Phrase with sentence beginning/ending tag

<s>	西藏	金融	工作	取得	显著	成绩	</s>
<s>	Tibet's	financial	Work	has gained	remarkable	achievements	</s>

- Motivation
 - Use tendency of Phrases to be located at the beginning or ending of sentence for their reordering
 - Tag the beginning/ending of sentence with <s>/</s>, borrowed from language modeling
- Training
 - Mark word-aligned sentences with beginning/ending tags
 - Treat beginning/ending tags as words
 - Extract bilingual phrases and MaxEnt-based reordering features
 - <s>西藏 ||| <s> Tibet's
- Decoding
 - Annotate source sentence with beginning/ending tags

Outline

- Brief introduction of BTG-based SMT
- Major issues of BTG-based SMT
- Proposed refinements for BTG-based SMT
- **Evaluation of proposed refinements**
- **Conclusions**

Evaluation

- Corpus and training
 - 2.4 M sentences from LDC, with 68.1M Chinese words and 73.8M English words
 - Four-gram language model using Xinhua section of the English Gigaword (181M words)
- Dev and test set
 - 580 sentences (not exceeding 50 characters) from NIST MT-02 test data
 - NIST MT-05 as the test set

Evaluation: reordering heuristics

	BLEU		Secs/sent	
BRUIN	29.96		54.3	
m	SW	SW+PR	SW	SW+PR
5	29.65	29.95	42.6	41.2
10	30.55	31.27	46.2	41.8
15	30.26	31.40	48.0	42.2
20	30.19	31.42	49.1	43.2

← increasing →

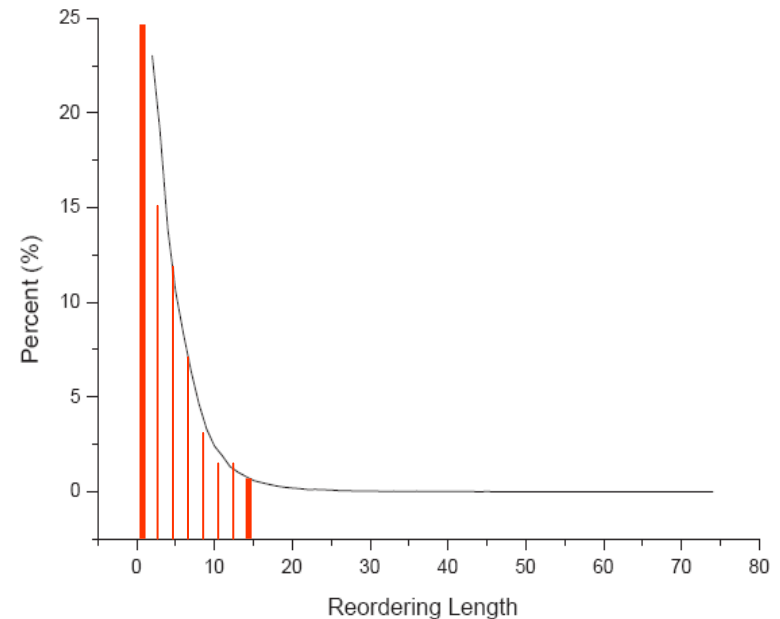
← decreasing →

Reordering heuristics have a two-fold effect on the performance:

- improving the BLEU score
- Decreasing the average decoding time

Swapping window size

- 15-word swapping window is the best choice for the tradeoff between decoding time and BLEU score
 - Reorderings within a window not exceeding 15 words have a very high proportion, 97.29%, in FBIS corpus.



Evaluation: phrases with special tags

	BLEU
Without special phrases	31.40
With special phrases	32.01

- Special phrases accounts for only 1.95% of the total phrase used
- They obtain an absolute improvement of 0.6% in the BLEU score

Conclusions

- Simple reordering heuristics improve BTG-based SMT
 - Swapping window
 - Punctuation restriction
 - Preventing incorrect swappings and reducing search space
- Special phrases with sentence beginning/ending tags improve reordering

Thanks: Q&A

- Brief introduction of BTG-based SMT
- Major issues of BTG-based SMT
- Proposed refinements for BTG-based SMT
- Evaluation of proposed refinements
- Conclusions