Improving Word Alignment for Statistical Machine Translation

Yanjun Ma

National Centre for Language Technology
School of Computing
Dublin City University

NCLT/CNGL MT Workshop

July 23, 2008

Outline

Word Alignment and SMT

Word alignment is a fundamental component for *all* SMT systems

- Word-based SMT
- Phrase-based SMT (PB-SMT): conditional v.s. joint model
- Formal syntax-based SMT: hiero-style v.s. ITG
- Syntax-based SMT: Tree-to-String, String-to-Tree, Tree-to-Tree

Approaches to word alignment

- Generative word alignment
- Discriminative word alignment
- Heuristics-based word alignment
- A mixture of above mentioned approaches

Challenges in Word Alignment

Evaluation

- Intrinsic evaluation on gold-standard: AER, F-measure
- Extrinsic evaluation on MT task: influence on BLEU

Further improvement in alignment quality

- GIZA++ can produce fairly good alignment

Correlation between word alignment quality and translation quality (cf. Patrik's talk)

- AER is shown not to correlate with BLEU, F-measure is better, but not good enough

Outline

Exploiting Word Segmentation and Syntax for Word Alignment

Word segmentation

- Question the *monolingual* segmentation of languages, i.e. Chinese

Syntax

- Make use of monolingual processing technology (parsing)

Word Packing [Ma et al., 2007]

Motivation

- *Monolingual* segmenters bring noise into alignment task
- 1-to-n alignment is one of the most difficult cases for word alignment

白葡萄酒: white wine	closest: 最 近
百货公司: department store	fifteen: 十 五
抱歉: excuse me	fine: 很 好
报警: call the police	flight: 次 航班
杯: cup of	get: 拿 到
必须: have to	here: 在 这里

Work Flow

- Use a word aligner to extract candidate 1–n alignments
  - Bilingual sentence
    我 想 要 一 杯 茶 。
    I 'd like a cup of tea .
  - Candidate extracted: $a_i =< c_i, E_i >$
    想 要 : 'd like
    杯 : cup of
- Estimate the reliability of the candidates $a_i =< c_i, E_i >$
  - Co-occurrence frequency: $COOC(c_i, E_i)$
  - Alignment confidence: $AC(a_i) = \frac{C(a_i)}{COOC(c_i, Ei)}$
- Pack the reliable candidates with a single token
- Re-iterate word alignment k times and stop when (i) there is no more packing to perform, or (ii) no improvement in translation is seen on the development set

Improved MT on IWSLT 2006 and IWSLT 2007 Chinese–English translation

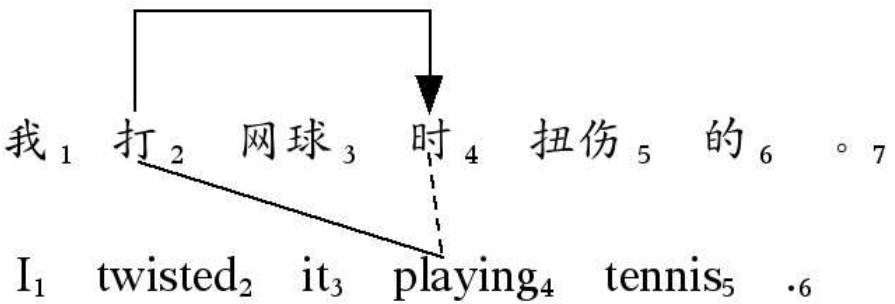
	IWSLT 2006	IWSLT 2007
Baseline	0.1855	0.2897
Iteration=1	0.1902	-
Iteration=2	0.1945	0.3000

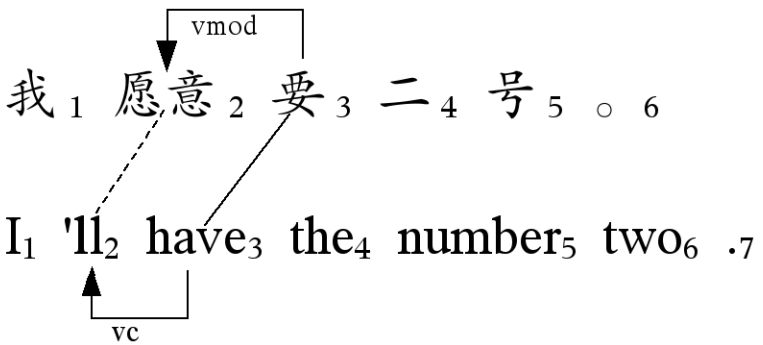
- Statistically significant ($p < 0.05$) improvement
- On a relatively small training data (40k sentence pairs)

Word Alignment using Syntactic Dependencies [Ma et al., 2008]

Motivation

- Incorporate syntactic dependencies into word alignment





General model

$$p(A|c_1^J, e_1^I) = \frac{1}{Z} \cdot p_\epsilon(A_\Delta|c_1^J, e_1^I) \cdot \prod_{j \in \bar{\Delta}} p(a_j|c_1^J, e_1^I, a_1^{j-1}, A_\Delta)$$

- Decomposed into an anchor alignment model and a syntax-enhanced model
- Syntax-enhanced model

$$\begin{aligned} \hat{a}_j &= \underset{a_j}{\text{argmax}}\{p_{\lambda_1^M}(a_j|c_1^J, e_1^I, a_1^{j-1}, A_\Delta)\} \\ &= \underset{a_j}{\text{argmax}}\{\sum_{m=1}^{M} \lambda_m h_m(c_1^J, e_1^I, a_1^j, A_\Delta, T_c, T_e)\} \end{aligned}$$

  - Assuming a set of anchor alignment can be obtained, the syntactic dependencies can be transformed into features in a discriminative word alignment model

Improved MT on IWSLT 2008 Chinese–English translation

	CHALLENGE	BTEC
Baseline	0.3194	0.3595
Syntax-Enhanced	0.3452	0.3823

- Statistically significant ($p < 0.05$) improvement
- Especially effective with a small training data (20k sentence pairs)

Outline

Word Alignment, MaTrEx and MT Evaluation

Word alignment modules in MaTrEx (cf. John's talk)

- Giza++, MTTK
- Word packing and its parallelisation
- Syntax-enhanced word alignment (not integrated)

Evaluation Campaign

- IWSLT 2007: Arabic–, Chinese–, Japanese–English [Hassan et al., 2007]
- IWSLT 2008: Arabic–, Chinese–English, English–Chinese

Outline

Further Work

Error analysis for word alignment

- GIZA++
- Word packing
- Word alignment with syntactic dependencies
- Different errors when optimised with different criteria (AER v.s. BLEU)

Word packing incorporating syntactic dependencies

- Direct packing using dependencies
- Incorporate dependency (packing) information into HMM alignment models
- Comparison with fertility-based models

Optimising word alignment for SMT

Open Research Topics

Word alignment and MT

- Improving generative word alignment models
- Improving discriminative word alignment models
- Impact of segmentation models on word alignment models
- Impact of syntactic information on word alignment models
- Gold-standard construction: guidelines for manual word alignment
- Evaluation of word alignment: intrinsic v.s. extrinsic evaluation
- Correlation between internal and external word alignment quality
- Optimising word alignment for MT
- Word alignment and EBMT
- Evaluating word alignment using other applications besides MT

Thanks for your attention

yma@computing.dcu.ie
http://www.computing.dcu.ie/~yma

References

Hassan, H., Ma, Y., and Way, A. (2007).
MaTrEx: the DCU machine translation system for IWSLT 2007.
In *Proceedings of the International Workshop on Spoken Language Translation*, pages 21–28, Trento, Italy.

Ma, Y., Ozdowska, S., Sun, Y., and Way, A. (2008).
Improving word alignment using syntactic dependencies.
In *Proceedings of the ACL-08: HLT Second Workshop on Syntax and Structure in Statistical Translation (SSST-2)*, pages 69–77, Columbus, OH.

Ma, Y., Stroppa, N., and Way, A. (2007).
Bootstrapping word alignment via word packing.
In *Proceedings of the 45th Annual Meeting of the Association of Computational Linguistics*, pages 304–311, Prague, Czech Republic.