Syntactic SMT Using a Discriminative Text Generation Model

Yue Zhang
SUTD, Singapore
yue_zhang@sutd.edu.sg

Kai Song
NEU, China
songkai.sk@alibaba-inc.com

Linfeng Song
ICT/CAS, China
songlinfeng@ict.ac.cn

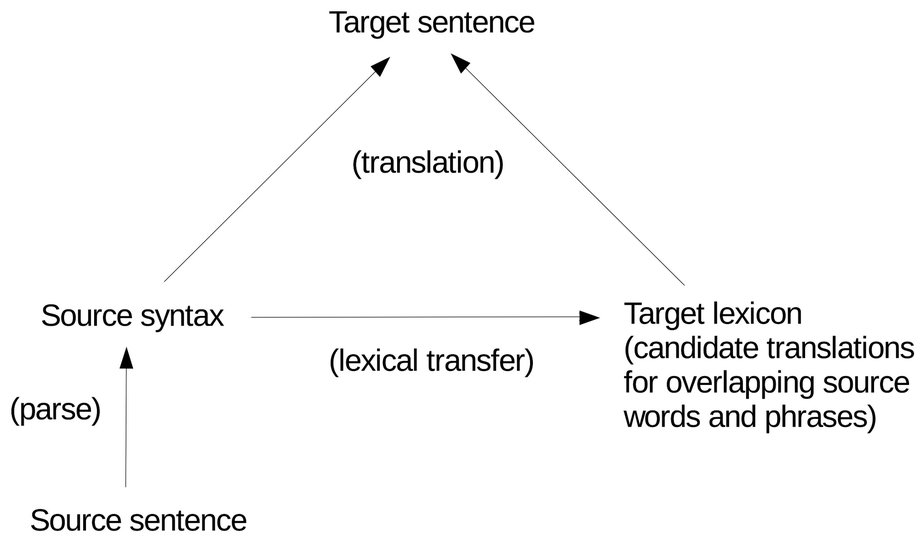
Jingbo Zhu
NEU, China
zhujingbo@mail.neu.edu.cn

Qun Liu
CNGL, Ireland
qliu@computing.dcu.ie

Introduction

- Syntactic Machine Translation
  - Translation by ***parsing*** (traditional)
    - Translation rules
    - Encode target order
    - Advantage:
      - Efficient
    - Disadvantage:
      - On adequacy:
        - noise, coverage, rules
      - On fluency:
        - no free ordering
  - Translation by ***generation*** (this work)
    - No translation rules
    - parse → transfer → synthesis
    - Advantage:
      - More psycho-linguistically motivated
      - No hard rules
      - Soft source constraints in target synthesis
    - Disadvantage:
      - complexity
- Our work
  - Preliminary study on translation by generation
  - Based on recent work: syntactic linearization (Zhang, 2013)
  - Perform word selection
  - Add bilingual features

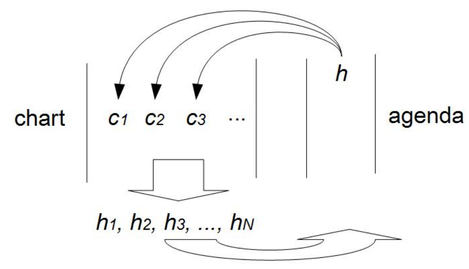
Approach



- Source parser: ZPar (Zhang and Nivre, 2011)
- Lexical transfer
  - IBM model 4 alignment
  - Consistent and cohesive phrase extracted
  - Target projective span
  - Filter translation options by probability
- Synthesis
  - Input: translation options
  - Mutually exclusive by source coverage
  - Output: target dependency tree

Synthesis

- Based on Zhang (2013)
- Search
  - Learning guided search
- Model
  - Scaled linear model



$$Score(e) = \frac{\vec{\theta} \cdot \Phi(e)}{|e|}$$

- Features
  - Base monolingual features

dependency syntax
$\text{WORD}(h) \cdot \text{POS}(h) \cdot \text{NORM}(size)$,
$\text{WORD}(h) \cdot \text{NORM}(size)$, $\text{POS}(h) \cdot \text{NORM}(size)$
$\text{POS}(h) \cdot \text{POS}(m) \cdot \text{POS}(b) \cdot dir$
$\text{POS}(h) \cdot \text{POS}(h_l) \cdot \text{POS}(m) \cdot \text{POS}(m_r) \cdot dir\ (h > m)$,
$\text{POS}(h) \cdot \text{POS}(h_r) \cdot \text{POS}(m) \cdot \text{POS}(m_l) \cdot dir\ (h < m)$
$\text{WORD}(h) \cdot \text{POS}(m) \cdot \text{POS}(m_l) \cdot dir$,
$\text{WORD}(h) \cdot \text{POS}(m) \cdot \text{POS}(m_r) \cdot dir$
$\text{POS}(h) \cdot \text{POS}(m) \cdot \text{POS}(m_1) \cdot dir$,
$\text{POS}(h) \cdot \text{POS}(m_1) \cdot dir$, $\text{POS}(m) \cdot \text{POS}(m_1) \cdot dir$
$\text{WORD}(h) \cdot \text{POS}(m) \cdot \text{POS}(m_1) \cdot \text{POS}(m_2) \cdot dir$,
$\text{POS}(h) \cdot \text{POS}(m) \cdot \text{POS}(m_1) \cdot \text{POS}(m_2) \cdot dir$,
...

dependency syntax for completed words
$\text{WORD}(h) \cdot \text{POS}(h) \cdot \text{WORD}(h_l) \cdot \text{POS}(h_l)$,
$\text{POS}(h) \cdot \text{POS}(h_l)$,
$\text{WORD}(h) \cdot \text{POS}(h) \cdot \text{POS}(h_l)$,
$\text{POS}(h) \cdot \text{WORD}(h_l) \cdot \text{POS}(h_l)$,
$\text{WORD}(h) \cdot \text{POS}(h) \cdot \text{WORD}(h_r) \cdot \text{POS}(h_r)$,
$\text{POS}(h) \cdot \text{POS}(h_r)$,
...

surface string patterns (B—bordering index)
$\text{WORD}(B-1) \cdot \text{WORD}(B)$, $\text{POS}(B-1) \cdot \text{POS}(B)$,
$\text{WORD}(B-1) \cdot \text{POS}(B)$, $\text{POS}(B-1) \cdot \text{WORD}(B)$,
$\text{WORD}(B-1) \cdot \text{WORD}(B) \cdot \text{WORD}(B+1)$,
$\text{WORD}(B-2) \cdot \text{WORD}(B-1) \cdot \text{WORD}(B)$,
$\text{POS}(B-1) \cdot \text{POS}(B) \cdot \text{POS}(B+1)$,
...

surface string patterns for complete sentences
$\text{WORD}(0)$, $\text{WORD}(0) \cdot \text{WORD}(1)$,
$\text{WORD}(size-1)$,
$\text{WORD}(size-1) \cdot \text{WORD}(size-2)$,
$\text{POS}(0)$, $\text{POS}(0) \cdot \text{POS}(1)$,
$\text{POS}(0) \cdot \text{POS}(1) \cdot \text{POS}(2)$,
...

  - New bilingual features

bilingual syntactic features ($\text{LEN}(path) \leq 3$)
$\text{POS}(th) \cdot \text{POS}(tm) \cdot dir \cdot \text{LABELS}(path)$,
$\text{WORD}(th) \cdot \text{POS}(tm) \cdot dir \cdot \text{LABELS}(path)$,
$\text{POS}(th) \cdot \text{WORD}(tm) \cdot dir \cdot \text{LABELS}(path)$,
$\text{WORD}(th) \cdot \text{WORD}(tm) \cdot dir \cdot \text{LABELS}(path)$,
$\text{WORD}(sh) \cdot \text{WORD}(sm) \cdot dir \cdot \text{LABELS}(path)$,
$\text{WORD}(sh) \cdot \text{WORD}(th) \cdot dir \cdot \text{LABELS}(path)$,
$\text{WORD}(sm) \cdot \text{WORD}(tm) \cdot dir \cdot \text{LABELS}(path)$,
$\text{POS}(th) \cdot \text{POS}(tm) \cdot dir \cdot \text{LABELSPOS}(path)$,
$\text{WORD}(th) \cdot \text{POS}(tm) \cdot dir \cdot \text{LABELSPOS}(path)$,
$\text{POS}(th) \cdot \text{WORD}(tm) \cdot dir \cdot \text{LABELSPOS}(path)$,
$\text{WORD}(th) \cdot \text{WORD}(tm) \cdot dir \cdot \text{LABELSPOS}(path)$,
$\text{WORD}(sh) \cdot \text{WORD}(sm) \cdot dir \cdot \text{LABELSPOS}(path)$,
$\text{WORD}(sh) \cdot \text{WORD}(th) \cdot dir \cdot \text{LABELSPOS}(path)$,
$\text{WORD}(sm) \cdot \text{WORD}(tm) \cdot dir \cdot \text{LABELSPOS}(path)$,

phrase translation features
$\text{PHRASE}(m) \cdot \text{PHRASE}(t), P(trans)$,

bilingual syntactic features
$\text{POS}(th) \cdot \text{POS}(tm) \cdot dir \cdot \text{LEN}(path)$,
$\text{WORD}(th) \cdot \text{POS}(tm) \cdot dir \cdot \text{LEN}(path)$,
$\text{POS}(th) \cdot \text{WORD}(tm) \cdot dir \cdot \text{LEN}(path)$,
$\text{WORD}(th) \cdot \text{WORD}(tm) \cdot dir \cdot \text{LEN}(path)$,
$\text{WORD}(sh) \cdot \text{WORD}(sm) \cdot dir \cdot \text{LEN}(path)$,
$\text{WORD}(sh) \cdot \text{WORD}(th) \cdot dir \cdot \text{LEN}(path)$,
$\text{WORD}(sm) \cdot \text{WORD}(tm) \cdot dir \cdot \text{LEN}(path)$,

Experiments

- Dataset
  - IWSLT 2010 Chinese-English
- Baselines (traditional syntactic)
  - String to tree (S2T)
  - Tree to string (T2S)
  - Tree to tree (T2T)
- Results

System	T2S	S2T	T2T	OURS
BLEU	32.65	36.07	28.46	34.24

References

- Yue Zhang, 2013. Partial-Tree Linearization: Generalized Word Ordering for Text Synthesis. In proceedings of IJCAI.
- Yue Zhang and Joakim Nivre, 2011. Transition-Based Dependency Parsing with Rich Non-Local Features. In proceedings of ACL.

Acknowledgement

The work has been supported by the Singapore Ministration of Education Tier 2 project T2MOE201301 and the startup grant SRG ISTD 2012 038 from SUTD.