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
CNGI, 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
    \[ \text{Score}(e) = \frac{\hat{\theta} \cdot \Phi(e)}{|e|} \]
- Features
  - Base monolingual features
  - New bilingual features

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 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.