Semantic Cohesion Model for Phrase-based SMT

Minwei Feng\textsuperscript{1}, Weiwei Sun\textsuperscript{2} and Hermann Ney\textsuperscript{1}

COLING 2012 December 13, 2012

Human Language Technology and Pattern Recognition\textsuperscript{1}
Lehrstuhl für Informatik 6
Computer Science Department
RWTH Aachen University, Aachen, Germany

The MOE Key Laboratory of Computational Linguistics\textsuperscript{2}
Institute of Computer Science and Technology
Peking University, Beijing, China
Outline

- Introduction
- Previous Work
- Semantic Cohesion model
- Comparison to Syntactic Cohesion model
- Experiments
- Conclusion
Introduction

- a novel semantic cohesion model
- use source-side predicate-argument structures as soft constraints
- designed for phrase-based MT
- added as a new feature in the log-linear framework
- experiments on Chinese-English NIST tasks
- 0.93 BLEU 0.98 TER improvements on average
- compare our method with [Cherry 08]
Previous Work

- **[Wu & Fung 09]**
  - A hybrid two-pass method to use semantics for SMT
  - Phrase-based SMT decoder + re-ordered translation hypothesis with the maximum match of semantic predicates and arguments
  - A semantics-based post reordering technique
  - Limited usage of semantic information, the second pass can only do some permutation of the output of the phrase-based decoder.

- **[Liu & Gildea 10]**
  - Two semantic role features in the tree-to-string MT system
  - First feature describes reordering of the source side semantic roles in the target side
  - Second feature penalizes the deletion of source side semantic roles
  - Improvement over a small FBIS English-to-Chinese system.
Previous Work

▸ [Gao & Vogel 11]
  ▶ utilizing target side SRL information to improve the hierarchical phrase-based machine translation
  ▶ extract SRL-aware SCFG rules together with conventional Hiero rules
  ▶ special conversion rules are applied during rule extraction procedure to ensure that when SRL-aware SCFG rules are used in derivation
  ▶ the decoder only generates hypotheses with complete semantic structures

▸ [Baker & Bloodgood+ 12]
  ▶ a modality/negation (MN) annotation scheme in the target side
  ▶ define modality as an extra-propositional component of meaning and negation as an inextricably intertwined component of modality
  ▶ A MN lexicon is created in a semi-supervised way
  ▶ a target side MN tagger to identify the substrings that are related to MN
  ▶ incorporate the MN into a small Urdu-English translation task
  ▶ using a tree grafting approach, the original syntactic tags enriched with new MN annotations are assigned to the parse trees of target side training data
Semantic Cohesion model

- Semantic Role Labeling (SRL)
- recognizing arguments involved by predicates in a given sentence and labeling their semantic types
- typical semantic classes include:
  - core arguments to a predicate: Agent, Patient, Source, Goal
  - adjuncts: Location, Time, Manner, Cause
- sentence-level SRL is concerned with the characterization of events
- understand the essential meaning of the original input sentences – who did what to whom, for whom or what, how, where, when and why?
- this shallow semantic interpretation abstracts important predicate-argument structural information away from syntactic structure and may potentially benefit machine translation, as well as many other NLP applications.
Bell, based in Los Angeles, makes and distributes electronic, computer and building products.

An example of SRL. The source sentence is given at top.

- four rows
- first row is original source sentence
- next three rows are event layers
- each event layer has one predicate and corresponding arguments
- predicate *make* and predicate *distribute* governs two arguments A0 (namely proto-Agent) and A1 (namely proto-Patient) respectively; predicate *base* possesses two semantic roles A1 and AM-LOC (namely location)
Semantic Cohesion Model: algorithm

Given the source sentence and its predicate-argument structural information, during the translation process, every time one hypothesis is extended, the added model checks if the source semantic analysis contains one structure $S$ such that:

- Its translation is already started (at least one word is covered)
- It is interrupted by the new added phrase (at least one word in the new source phrase is not in $S$)
- It is not finished (after the new phrase is added, there is still at least one uncovered source word in $S$)

If so, we say this hypothesis violates the structure $S$, and the model returns the number of structures that this hypothesis violates. We use two kinds of structures; we add two models/features into the log-linear framework.
Semantic Cohesion Model: algorithm

Bell, based in Los Angeles, makes and distributes electronic, computer and building products.

An example of SRL. The source sentence is given at top.

- two kinds of structures
  - the whole predicate-argument, i.e. event layer (SRL1)
    how many event layers that one search state violates
  - semantic role (SRL2)
    how many semantic roles that one search state violates

- one example
  - only *Bell* is already translated and the decoder decides to translate *computer*
    - this decision violates the third event layer; feature value of SRL1 is 1
    - this decision violates two semantic role A0; feature value of SRL2 is 2
A Semantic Cohesion Model: comparison

Comparison with Syntactic cohesion model [Cherry 08]

[ Bell, based in Los Angeles, makes and distributes electronic, computer and building products. ]

A dependency tree example. The source sentence is given at bottom.
A Semantic Cohesion Model: comparison

Comparison with Syntactic Cohesion model [Cherry 08]

Given the source sentence and its dependency tree, during the translation process, once a hypothesis is extended, check if the source dependency tree contains a subtree $T$ such that:

- Its translation is already started (at least one node is covered)
- It is interrupted by the new added phrase (at least one word in the new source phrase is not in $T$)
- It is not finished (after the new phrase is added, there is still at least one free node in $T$)

If so, we say this hypothesis violates the subtree $T$, and the model returns the number of subtrees that this hypothesis violates.
[Bell, based in Los Angeles, makes and distributes electronic, computer and building products.]

- filled nodes means the words have been already translated
- suppose the length of the new added phrase is 1
- only position 5 and 6 (green rectangle) are good candidates
- choosing other source words (red rectangle) to translate will violate the subtree $in_4 - Los_5 - Angeles_6$
A Semantic Cohesion Model: experimental setup

- Lowercased training data from BOLT task alignment trained with GIZA++
- Tuning corpus: NIST06 test corpora: NIST02 03 04 05 and 08
- 5-gram LM (1,694,412,027 running words) trained by SRILM toolkit [Stolcke 02] with modified Kneser-Ney smoothing
  LM training data: target side of bilingual data.
- BLEU [Papineni & Roukos+ 01] and TER [Snover & Dorr+ 05] reported all scores calculated in lowercase way.
- Stanford Parser [Levy & Manning 03] used to get the Chinese constituent tree for the SRL and the dependency tree for the syntactic cohesion model

<table>
<thead>
<tr>
<th></th>
<th>Chinese</th>
<th>English</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sentences</td>
<td>5,384,856</td>
<td></td>
</tr>
<tr>
<td>Running Words</td>
<td>115,172,748</td>
<td>129,820,318</td>
</tr>
<tr>
<td>Vocabulary</td>
<td>1,125,437</td>
<td>739,251</td>
</tr>
</tbody>
</table>

training data statistics
### A Semantic Cohesion Model: results

<table>
<thead>
<tr>
<th>Systems</th>
<th>NIST02</th>
<th>NIST03</th>
<th>NIST04</th>
<th>NIST05</th>
<th>NIST08</th>
<th>avg.</th>
<th>Index</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>BLEU scores</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>baseline</td>
<td>33.60</td>
<td>34.29</td>
<td>35.73</td>
<td>32.15</td>
<td>26.34</td>
<td>-</td>
<td>1</td>
</tr>
<tr>
<td>baseline+SRL1</td>
<td>34.50</td>
<td>34.76</td>
<td>36.21</td>
<td>32.75</td>
<td>26.77</td>
<td>0.57</td>
<td>2</td>
</tr>
<tr>
<td>baseline+SRL2</td>
<td>34.73</td>
<td>34.88</td>
<td>36.60</td>
<td>32.83</td>
<td>26.82</td>
<td>0.75</td>
<td>3</td>
</tr>
<tr>
<td>baseline+SRL1+SRL2</td>
<td>35.05</td>
<td>34.93</td>
<td>36.71</td>
<td>33.22</td>
<td>26.89</td>
<td>0.93</td>
<td>4</td>
</tr>
<tr>
<td>baseline+SC</td>
<td>34.96</td>
<td>34.52</td>
<td>36.37</td>
<td>33.35</td>
<td>26.90</td>
<td>0.79</td>
<td>5</td>
</tr>
<tr>
<td><strong>TER scores</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>baseline</td>
<td>61.36</td>
<td>60.48</td>
<td>59.12</td>
<td>60.94</td>
<td>65.17</td>
<td>-</td>
<td>6</td>
</tr>
<tr>
<td>baseline+SRL1</td>
<td>60.44</td>
<td>59.58</td>
<td>58.61</td>
<td>60.01</td>
<td>64.88</td>
<td>0.71</td>
<td>7</td>
</tr>
<tr>
<td>baseline+SRL2</td>
<td>60.28</td>
<td>59.49</td>
<td>58.14</td>
<td>60.20</td>
<td>65.11</td>
<td>0.77</td>
<td>8</td>
</tr>
<tr>
<td>baseline+SRL1+SRL2</td>
<td>60.05</td>
<td>59.55</td>
<td>58.14</td>
<td>59.69</td>
<td>64.74</td>
<td>0.98</td>
<td>9</td>
</tr>
<tr>
<td>baseline+SC</td>
<td>59.90</td>
<td>59.37</td>
<td>58.27</td>
<td>59.69</td>
<td>64.44</td>
<td>1.08</td>
<td>10</td>
</tr>
</tbody>
</table>

SC: [Cherry 08] how many subtrees that one search state violates  
SRL1: how many event layers that one search state violates  
SRL2: how many semantic roles that one search state violates
A Semantic Cohesion Model: conclusions

- proposed a method to utilize the source side semantic analysis for SMT
- two features
  - SRL1: global semantic coherence
  - SRL2: local semantic coherence
  - SRL1: long distance reordering model
  - SRL2: local reordering model/phrase model (it encourages the decoder to choose those phrases that keep semantic coherence)
- semantic analysis information improves the baseline 0.93 BLEU and 0.98 TER
- comparative study with Syntactic Cohesion model [Cherry 08]
  - SRL is 0.14 BLEU better than SC while SC is 0.1 TER better than SRL.
  - compared to full parsing model SC: our model abstracts important event structures away from syntactic parses our model concentrates on modeling the skeleton of a sentence and provide much less information experiments suggest that SRL achieves an equivalent contribution (as constraints) to a phrase-based MT system
Thank you for your attention

Minwei Feng

feng@i6.informatik.rwth-aachen.de

http://www-i6.informatik.rwth-aachen.de/
References


Linguistics, COLING ’10, pp. 716–724, Stroudsburg, PA, USA, 2010. Association for Computational Linguistics. 4


