A Few Thoughts and Attempts on the Future of SMT

Deyi Xiong

Soochow University, Suzhou, China
Tower of Humankind
Tower of Machine Translation

- Source words
- Source syntax
- Source semantics
- Interlingual
- Target semantics
- Target syntax
- Target words
Tower of Machine Translation

Where are we?
Tower of Machine Translation

Where are we?
Where are we?

Where will we be?
SMT: Future Directions

- Representation
- Unsupervised learning
- Semantics, discourse and SMT
Representation

- Knowledge representation
  - one of central concepts in AI

- Translation representation
  - word pairs → word-based SMT
  - bilingual phrases → phrase-based SMT
  - ITG rules → ITG-based SMT
  - various tree-based rules → tree-based SMT
Representation

What’s next?
- distributional representation
- semantics-based whole-sentence representation
- other new representations …

Tradeoff
- computational tractability
- expressive capacity
Rule representations with topic distributions: **Xiao et al., 2012**

(a) 作战 能力 ⇒ operational capability

(b) 给予 $X_1$ ⇒ grants $X_1$

(c) 给予 $X_1$ ⇒ give $X_1$

(d) $X_1$ 举行 会谈 $X_2$ ⇒ held talks $X_1$ $X_2$
Topic Similarity Model

Source Document

授予 X1 ⇔ grants X1

授予 X1 ⇔ give X1
Topic Similarity Model: Bias

- Flat distribution is dissimilar to sharp distribution.
- Topic similarity model: **punish** rules with even distributions.
  - Topic-insensitive rule
Topic Sensitivity Model

- Calculate entropy of rule-topic distributions of rules
- Compensate the bias of the topic similarity model
SMT: Future Directions

- Representation
- Unsupervised learning
- Semantics, discourse and SMT
Unsupervised Learning

- EM for word alignment: big success
- EM for phrase/grammar induction: failure or moderate success
- Unsupervised discriminative learning: an exciting direction for grammar induction (Xiao et al., 2012; Xiao and Xiong 2013)
Max-Margin Synchronous Grammar Induction: Xiao and Xiong 2013

奥巴马 与 内塔尼亚胡 举行 会谈

Obama hold a talk with Netanyahu

奥巴马 $X_1 \rightarrow$ Obama $X_1$

与 内塔尼亚胡 $\rightarrow$ with Netanyahu

$X_1$ 举行 会谈 $\rightarrow$ hold a talk $X_1$

......
Max-Margin SGI

Max-margin

I - BLEU

Non-local features

Target Parse Structures

Obama hold a talk with Neta.

Obama 与 内塔. 举行 会谈

X

X

X

X

X

X

X
Optimization
For each sent.

1. Biparse Reference
2. Collect Rules
3. Cost-Augmented Viterbi
4. Biparse Viterbi
5. Update Weights

2014/5/17 YSSNLP 2014
Optimization

For each sent.

Biparse Reference

Collect Rules

Cost-Augmented Viterbi

Biparse Viterbi

Update Weights

Obama hold a talk with Neta.
Optimization

For each sent.

Biparse Reference

Collect Rule

Cost-Augmented Viterbi

Biparse Viterbi

Update Weights

Obama hold a talk with Neta.

X_1 举行 会谈 → hold a talk X_1
Optimization

For each sent.

Biparse Reference

Collect Rules

Cost-Augmented Viterbi

Biparse Viterbi

Update Weights

\[ f(s^{(i)}, t) \rightarrow \text{BLUE-4}(t^{(i)}, t) \]
Optimization

For each sent.

Biparse Reference

Collect Rules

Cost-Augmented Viterbi

Biparse Viterbi

Update Weights

Obama with Neta. hold a talk

0,1 0,1 1,3 1,3 3,5 3,6

0,5 0,6

0,3 0,3

1,5 1,6

2014/5/17

YSSNLP 2014
Optimization

For each sent.

1. Sub-gradient: reference - viterbi
2. Projection: rescale weights

Shalev shwartz et al. (2007)
Future Unsupervised Learning for SMT

- Learning from big and unstructured data
- Learning tailored for new representations
- Never-ending learning
SMT: Future Directions

- Representation
- Unsupervised learning
- Semantics, discourse and SMT
Tower of Machine Translation

- Source words
  - Source syntax
    - Interlingual
    - Target semantics
  - Source semantics
- Target words
  - Target syntax
Semantics and SMT: Multi-Level View

- **Lexical semantics**
  - word senses (Xiong and Zhang, 2014)
  - semantic roles (Xiong et al., 2012)

- **Sentential semantics**
  - compositional semantics

- **Discourse-level semantics**
  - cohesion (Xiong et al., 2013b, 2013c)
  - coherence (Xiong et al., 2013a)
Semantics and SMT: Multi-Level View

- **Lexical semantics**
  - word senses (Xiong and Zhang, 2014)
  - semantic roles (Xiong et al., 2012)

- **Sentential semantics**
  - compositional semantics

- **Discourse-level semantics**
  - cohesion (Xiong et al., 2013b, 2013c)
  - coherence (Xiong et al., 2013a)
Hidden Word Senses for SMT: Xiong and Zhang 2014

More info: A Sense-Based Translation Model for SMT

<table>
<thead>
<tr>
<th>source sentences</th>
<th>HDP-based WSI</th>
<th>sense-tagged source sentences</th>
<th>MaxEnt classifiers</th>
<th>decoder</th>
<th>sense-based translation model</th>
<th>target sentences</th>
</tr>
</thead>
</table>

Other models

Baseline

STM (sense)

STM (sense+lexicon)

BLEU-4

<table>
<thead>
<tr>
<th>s1</th>
<th>s2</th>
<th>s3</th>
</tr>
</thead>
<tbody>
<tr>
<td>运管 (operate)</td>
<td>运营 (operate)</td>
<td>运管 (operate)</td>
</tr>
<tr>
<td>设施 (facility)</td>
<td>卫星 (satellite)</td>
<td>市场 (market)</td>
</tr>
<tr>
<td>计划 (plan)</td>
<td>系统 (system)</td>
<td>企业 (enterprise)</td>
</tr>
<tr>
<td>基础 (foundation)</td>
<td>国家 (country)</td>
<td>竞争 (competition)</td>
</tr>
<tr>
<td>项目 (project)</td>
<td>提供 (supply)</td>
<td>资产 (assets)</td>
</tr>
<tr>
<td>公司 (company)</td>
<td>国际 (inter-nation)</td>
<td>利润 (profit)</td>
</tr>
<tr>
<td>结构 (structure)</td>
<td>机构 (institution)</td>
<td>造成 (cause)</td>
</tr>
<tr>
<td>服务 (service)</td>
<td>进行 (proceed)</td>
<td>费用 (cost)</td>
</tr>
<tr>
<td>组织 (organization)</td>
<td>中心 (center)</td>
<td>资金 (capital)</td>
</tr>
<tr>
<td>提供 (supply)</td>
<td>合作 (cooperate)</td>
<td>业务 (business)</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>s4</th>
<th>s5</th>
</tr>
</thead>
<tbody>
<tr>
<td>费用 (cost)</td>
<td>城市 (city)</td>
</tr>
<tr>
<td>股价 (share price)</td>
<td>处理 (process)</td>
</tr>
<tr>
<td>27000</td>
<td>自来水 (tap-water)</td>
</tr>
<tr>
<td>科索沃 (Kosovo)</td>
<td>工厂 (factory)</td>
</tr>
<tr>
<td>额外 (extra)</td>
<td>汽车 (car)</td>
</tr>
<tr>
<td>工资 (wage)</td>
<td>铁路 (railway)</td>
</tr>
<tr>
<td>美元 (dollar)</td>
<td>污水 (sewage)</td>
</tr>
<tr>
<td>商业 (commerce)</td>
<td>办公室 (office)</td>
</tr>
<tr>
<td>收入 (income)</td>
<td>保本 (break-even)</td>
</tr>
<tr>
<td>铁路局 (railway administration)</td>
<td>部件 (component)</td>
</tr>
<tr>
<td>处于 (lie)</td>
<td>拍照 (photograph)</td>
</tr>
<tr>
<td>DPRK</td>
<td>119</td>
</tr>
<tr>
<td>保险 (insurance)</td>
<td>超支 (overspend)</td>
</tr>
<tr>
<td>地位 (position)</td>
<td>经济 (economy)</td>
</tr>
<tr>
<td>竞争者 (competitor)</td>
<td>平衡 (balance)</td>
</tr>
</tbody>
</table>
Semantics and SMT: Multi-Level View

- Lexical semantics
  - word senses (Xiong and Zhang, 2014)
  - semantic roles (Xiong et al., 2012)

- Sentential semantics
  - compositional semantics

- Discourse-level semantics
  - cohesion (Xiong et al., 2013b, 2013c)
  - coherence (Xiong et al., 2013a)
Semantic Roles for SMT:
Xiong et al., 2012

**Predicate Translation model**

**Argument reordering model**

Decoder

<table>
<thead>
<tr>
<th>Reordering Category</th>
<th>Percent</th>
</tr>
</thead>
<tbody>
<tr>
<td>NC</td>
<td>82.43%</td>
</tr>
<tr>
<td>L2R</td>
<td>11.19%</td>
</tr>
<tr>
<td>R2L</td>
<td>6.38%</td>
</tr>
</tbody>
</table>

**Argument reordering category:**
- **NC:** no change across languages
- **L2R:** moving from the left side of its predicate to the right side after translation
- **R2L:** moving from the right side of predicate to the left side
Semantics and SMT: Multi-Level View

- Lexical semantics
  - word senses (Xiong and Zhang, 2014)
  - semantic roles (Xiong et al., 2012)

- Sentential semantics
  - compositional semantics

- Discourse-level semantics
  - cohesion (Xiong et al., 2013b, 2013c)
  - coherence (Xiong et al., 2013a)
Discourse-Level Cohesion: Xiong et al., 2013b

- **Direct reward model**: rewards translation hypotheses whenever lexical cohesion items occur in them.

- **Conditional probability model**: measures the appropriateness of using lexical cohesion items.

- **Mutual information trigger model**: estimates the strength of association between items in a lexical cohesion relation.
Discourse-Level Cohesion: Xiong et al., 2013c

Source chains generation
Compute lexical chains for each source document to be translated.

Target chains projection
Project the computed source lexical chains onto the corresponding target document via maximum entropy classifiers.

Cohesion model incorporation
Incorporate lexical cohesion into the target document translation via cohesion models built on the projected target lexical chains.
Semantics and SMT: Multi-Level View

- **Lexical semantics**
  - word senses (Xiong and Zhang, 2014)
  - semantic roles (Xiong et al., 2012)

- **Sentential semantics**
  - compositional semantics

- **Discourse-level semantics**
  - cohesion (Xiong et al., 2013b, 2013c)
  - coherence (Xiong et al., 2013a)
Discourse-Level Coherence:
Xiong and Zhang, 2013

- Model the sense continuity as a continuous sentence topic transition
- Characterize a document as a coherence chain of sentence topics
- Project source coherence chain to target documents
Welcome to our ACL 2014 Tutorial

Semantics, Discourse and SMT
Tutorial Outline

- SMT Overall Review (30 minutes)
  - SMT architecture
  - phrase- and syntax-based SMT

- Semantics and SMT (1 hour and 15 minutes)
  - Brief introduction of semantics
  - Lexical semantics for SMT
  - Semantic representations in SMT
  - Semantically Motivated Evaluation
  - Advanced topics: deep semantic learning for SMT
  - Future directions

- Discourse and SMT (1 hour and 15 minutes)
  - Introduction of discourse: linguistics, computational and bilingual discourse
  - Discourse-based SMT: modeling, training, decoding and evaluation
  - Future directions
Special thanks to:

Xinyan Xiao
Guosheng Ben
Yang Ding
References


References


Thank you!