Integrating a Rule-Based with a Hierarchical Translation System

Sarah Taylor

University of Tübingen

17 July 2018
1 Introduction
2 Previous Work
3 Architecture
4 Components
5 Experiments
6 Results
7 Conclusion
8 References
Warm Up Question

What are some of the advantages and disadvantages of RBMT and SMT systems, respectively?
Warm Up Question

What are some of the advantages and disadvantages of RBMT and SMT systems, respectively?

**RBMT**

- Hand-written rules domain independent, errors more easily correctable
- Don’t need bilingual texts, higher reusability
- However: writing rules and dictionaries costs effort, expensive

**SMT**

- Efficient use of data resources, language flexible
- However: statistical anomalies, new words, word and sentence alignment i.e. language word order and grammaticality tricky
RBMT and SMT Systems

- Rule-based systems have better results than statistical machine translation
- Hybrid system lets RBMT fill gaps in SMT lexical knowledge

Chen, Eisele 2010[1]
What differences do you see in the various translations? Which is best?

**Figure**: Translation examples (Chen, Eisele 2010[1])

<table>
<thead>
<tr>
<th>In-domain</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Source</strong></td>
</tr>
<tr>
<td><strong>Reference</strong></td>
</tr>
<tr>
<td><strong>Lucy</strong></td>
</tr>
<tr>
<td><strong>Moses</strong></td>
</tr>
<tr>
<td><strong>+Lucy</strong></td>
</tr>
<tr>
<td><strong>Joshua</strong></td>
</tr>
<tr>
<td><strong>+Lucy</strong></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Out-of-domain</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Source</strong></td>
</tr>
<tr>
<td><strong>Reference</strong></td>
</tr>
<tr>
<td><strong>Lucy</strong></td>
</tr>
<tr>
<td><strong>Moses</strong></td>
</tr>
<tr>
<td><strong>+Lucy</strong></td>
</tr>
<tr>
<td><strong>Joshua</strong></td>
</tr>
<tr>
<td><strong>+Lucy</strong></td>
</tr>
</tbody>
</table>
Eisele 2008 vs. Chen and Eisele 2010

- The core SMT system of Eisele et. al 2008 is replaced with a hierarchical phrase-based SMT system (HPBSMT) from Chiang 2007
- Aim to preserve syntax, add lexical information from RBMT, improve the grammaticality of translations
- Unrealistic in practice to use 6 plus RBMT systems plus the SMT core, have restricted here to 1 RBMT and SMT decoder here

Chen, Eisele 2010[1]
Phrase-based models

Phrase-based models can perform good translations, yet reordering not always present or successful:

![Phrase-based-IBM caption Alignment Template System and IBM phrase-based system (Chiang 2007)](image)

What needs improving here in the translation?
How to capitalize on strengths of phrase-based approach?
Warm Up HPBSMT

How does Chiang combines the strengths of phrase-based and syntax-based translation?
Chiang’s hierarchical phrase-based SMT system

- Phrases are good for learning the (re)ordering of words, can use it to learn this for phrases as well
- For this they use hierarchical phrases that can contain other phrases
- E.g. capturing how Chinese PPs usually modify VPs on left, whereas English on the right

Chiang 2007[2]
Figure: Hierarchical phrase pairs (Chiang 2007[2])

Figure: Hierarchical phrase-based SMT (Chiang 2007[2])
Introduction: SCFG

Synchronus context-free-grammar (SCFG)

$X \rightarrow \langle \gamma, \alpha, \sim \rangle$

- Also known as syntax-directed transduction grammar
- In synchronous CFG, elementary structures are re-write rules

Figure: SCFG and conventional phrase pairs (Chiang 2007[2])
Introduction: SCFG

Synchronous context-free grammars

- A type of formal grammar for use in transfer-based machine translation
- Rules in these grammars apply to two languages at the same time, capturing grammatical structures that are each other’s translations
- Rules specify the structure of two phrases at the same time; one in the source language and one in the target language
- Numeric indices indicate correspondences between non-terminals in both constituent trees

\[ X \rightarrow \langle \text{den } X_1 \text{ habe ich } X_2, \text{ I have } X_2 X_1 \rangle. \]

Which rules are generalized? (GER->ENG)
"
\text{"den Termin habe ich vergessen"}
"

Figure: (Chen, Eisele 2010\[1\])
Reflecting on past work

- Best method is trying to select best output from a number of systems so as to form a multi-engine system from a group of independent systems.
- Important to find the correspondences between alternatives proposed by different MT systems.
- Because of this, specialized module for word alignment is needed.
- Need to understand relation between building blocks and select best combination.

Chen, Eisele 2010^1]
Architecture Considerations

Figure: Overview of Possible Hybrid Structures (Eisele 2008 Presentation Euromatrix[4])
Chen and Eisele’s New System

- Uses an SMT system with a modified translation model
- Standard phrase-table augmented using entries from rule-based MT system
- SMT decoder recombines building blocks from RBMT and SMT system using standard algorithm

Chen, Eisele 2010[1]
Figure: Architecture of New System (Chen, Eisele 2010[1])
Components: SMT Decoder

Moses Toolkit and phrase-based SMT decoder

- Parsing-based decoder which assumes a probabilistic synchronous context-free grammar (SCFG)
- Only hierarchical rules considered, as in Chang 2007
- Decoder uses chart-parsing algorithm

Moses Toolkit

- Open-source toolkit for SMT
- Support for linguistically motivated factors
- Confusion network decoding; allows for translation of ambiguous input
- Efficient phrase translation table representation; uses prefix tree structure for source words and on demand loading

Koehn et. al 2007[5]
Chen, Eisele 2010[1]
Components: Core SMT System

Joshua

- A general-purpose, open source toolkit for parsing-based machine translation
- Similar to what Moses does for phrase-based machine translation
- Used to build the core SMT system

Li 2009[6]
Chen, Eisele 2010[1]
Components: RBMT System

**Lucy**
- A commercial translation engine with a rule-based (transfer-based) system
- System contains monolingual and bilingual lexicons, analysis grammars, transfer rules, generation module, etc
- Used as a black box here
- One of the best systems for English and German translation

Chen, Eisele 2010[1]
Components: RBMT System

Figure: Lucy Architecture (LucySoftware.com[Web:4])
Aligning the Elements

- Phrase table generated by RBMT system, here with gaps modeled as a SCFG
- No word alignment information from RBMT system
- Two steps of phrase table construction:
  1. Word alignment: final word alignment determined on result of grow-diag-final-and heuristic
  2. Phrase extraction: use a bilingual corpus with input texts and translation from RBMT system
- Union of RBMT model and statistical model from training corpus forms combined phrase table for final translations

Chen, Eisele 2010[1]
Components: Alignment

Figure: Example file generated by word alignment process (Moses[^1])

[^1]: Moses
## Components: Bilingual Corpus

<table>
<thead>
<tr>
<th>Parallel Corpus (L1-L2)</th>
<th>Sentences</th>
<th>L1 Words</th>
<th>English Words</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bulgarian-English</td>
<td>406,934</td>
<td></td>
<td>9,886,931</td>
</tr>
<tr>
<td>Czech-English</td>
<td>646,605</td>
<td>12,999,453</td>
<td>15,625,264</td>
</tr>
<tr>
<td>Danish-English</td>
<td>1,968,800</td>
<td>44,654,417</td>
<td>48,574,988</td>
</tr>
<tr>
<td>German-English</td>
<td>1,920,209</td>
<td>44,348,491</td>
<td>47,818,827</td>
</tr>
<tr>
<td>Greek-English</td>
<td>1,235,976</td>
<td></td>
<td>31,929,703</td>
</tr>
<tr>
<td>Spanish-English</td>
<td>1,965,734</td>
<td>51,575,748</td>
<td>49,093,806</td>
</tr>
<tr>
<td>Estonian-English</td>
<td>651,746</td>
<td>11,214,221</td>
<td>15,685,733</td>
</tr>
<tr>
<td>Finnish-English</td>
<td>1,924,942</td>
<td>32,266,343</td>
<td>47,460,063</td>
</tr>
<tr>
<td>French-English</td>
<td>2,007,723</td>
<td>51,388,643</td>
<td>50,196,035</td>
</tr>
<tr>
<td>Hungarian-English</td>
<td>624,934</td>
<td>12,420,276</td>
<td>15,096,358</td>
</tr>
<tr>
<td>Italian-English</td>
<td>1,909,115</td>
<td>47,402,927</td>
<td>49,666,692</td>
</tr>
<tr>
<td>Lithuanian-English</td>
<td>653,146</td>
<td>11,294,690</td>
<td>15,341,983</td>
</tr>
<tr>
<td>Latvian-English</td>
<td>637,599</td>
<td>11,928,716</td>
<td>15,411,680</td>
</tr>
<tr>
<td>Dutch-English</td>
<td>1,997,775</td>
<td>50,602,994</td>
<td>49,469,373</td>
</tr>
<tr>
<td>Polish-English</td>
<td>632,555</td>
<td>12,815,544</td>
<td>15,268,824</td>
</tr>
<tr>
<td>Portuguese-English</td>
<td>1,960,407</td>
<td>49,147,826</td>
<td>49,216,896</td>
</tr>
<tr>
<td>Romanian-English</td>
<td>399,375</td>
<td>9,628,010</td>
<td>9,710,331</td>
</tr>
<tr>
<td>Slovak-English</td>
<td>640,715</td>
<td>12,942,434</td>
<td>15,442,233</td>
</tr>
<tr>
<td>Slovene-English</td>
<td>623,490</td>
<td>12,525,644</td>
<td>15,021,497</td>
</tr>
<tr>
<td>Swedish-English</td>
<td>1,862,234</td>
<td>41,508,712</td>
<td>45,703,795</td>
</tr>
</tbody>
</table>

**Figure:** Sizes for parallel corpora after sentence aligning and removing XML (StatMT.org[Web:3])
Components: In Sum

Step by Step

- Phrase table from RBMT outputs with phrases with gaps, modeled as SCFG
- Phrase table construction: word alignment and phrase extraction
- RBMT phrase table made using bilingual corpus
- Alignment starts with mapping vocabularies in the RBMT corpus to those corresponding to the base model and initiating estimates
- Alignment includes diagonal neighboring words to existing aligned-points for better coverage
- Overall: the RBMT model and the statistical model from training corpus makes up the phrase table for final translations
- Hybrid system retains 6 features in total; risky but useful

Chen, Eisele 2010\[1\]
Experiments

Test Details

- In-domain and out-of-domain tasks
- Use of Europarl corpus
- Tested hybrid system using two test sets from WMT 2008 Shared Task
- Core SMT system with Joshua
- Hierarchical models trained on sentences with less than 80 tokens
- Statistical system includes a 5-gram language model that was constructed on the target side of the parallel corpus using SRILM toolkit
- Extracted small hierarchical model from Lucy’s translation of the development set and merged into large one

Chen, Eisele 2010[1]
Experiments

ZMERT
- Part of larger effort to develop Joshua into a software packing including hierarchical phrase-based decoders
- Publically available and functions as standalone application

Berkeley Aligner
- A software package which aligns words in multilingual parallel texts
- Implements recent innovations in unsupervised word alignment

Zaidan 2009[7]
CS Department UC Berkeley[Web:5]
Experimental Tools

- Using Joshua, it is a good idea to use ZMERT rather than other implementations for minimum error rate training.
- ZMERT determines feature weights and development sets to maximize BLEU score.
- Used Berkeley Aligner to align training data.
- For comparison, built another hybrid system with phrase-based SMT core using Moses Toolkit.
- For this, used the same data sets (training, tuning and testing), identical word alignments, the same language model.

Chen, Eisele 2010[1]
### Experiments

#### Figure: Example entries from combined phrase table (Chen, Eisele 2010[1])

<table>
<thead>
<tr>
<th>source</th>
<th>target</th>
<th>SMT features</th>
<th>RBMT features</th>
</tr>
</thead>
<tbody>
<tr>
<td>zum</td>
<td>at the</td>
<td>1.9800 1.8958 2.4356</td>
<td>1.9542 1.8255 2.1297</td>
</tr>
<tr>
<td>der X₁, die</td>
<td>the X₁ which</td>
<td>1.2552 1.7833 1.6795</td>
<td>1.0543 1.4845 1.4218</td>
</tr>
<tr>
<td>der X₁ der X₂</td>
<td>of the X₁ of the X₂</td>
<td>1.3979 1.1264 1.8677</td>
<td>1.58546 1.0686 1.5023</td>
</tr>
<tr>
<td>landesgrenzen</td>
<td>boundaries</td>
<td>1.1563 1.7584 1.1139</td>
<td>1.0 1.0 1.0</td>
</tr>
<tr>
<td>X₁ abgeschlossen sein</td>
<td>X₁ be finalised</td>
<td>1.8450 1.7077 1.8586</td>
<td>1.0 1.0 1.0</td>
</tr>
<tr>
<td>fakten X₁ der X₂</td>
<td>facts X₁ against the X₂</td>
<td>1.0413 1.0455 3.613</td>
<td>1.0 1.0 1.0</td>
</tr>
<tr>
<td>nach den</td>
<td>after that</td>
<td>1.0 1.0 1.0</td>
<td>1.1139 2.1035 2.129</td>
</tr>
<tr>
<td>auf der X₁</td>
<td>on which X₁</td>
<td>1.0 1.0 1.0</td>
<td>1.3617 1.4243 2.1300</td>
</tr>
<tr>
<td>die X₁ von X₂</td>
<td>who X₁ of X₂</td>
<td>1.0 1.0 1.0</td>
<td>1.3802 1.2750 1.9222</td>
</tr>
</tbody>
</table>
### Results

**Figure:** BLEU scores of German-English translations (Chen, Eisele 2010[1])

<table>
<thead>
<tr>
<th></th>
<th>In-domain</th>
<th>Out-of-domain</th>
</tr>
</thead>
<tbody>
<tr>
<td>Lucy</td>
<td>16.40</td>
<td>17.02</td>
</tr>
<tr>
<td>Moses</td>
<td>27.55</td>
<td>16.66</td>
</tr>
<tr>
<td>Moses+Lucy</td>
<td>27.26</td>
<td>16.06</td>
</tr>
<tr>
<td>Joshua</td>
<td>27.51</td>
<td>16.24</td>
</tr>
<tr>
<td>Joshua+Lucy</td>
<td>27.52</td>
<td>17.69</td>
</tr>
</tbody>
</table>

Results

What was seen?

- Hybrid system of Joshua and Lucy better than both in isolation, and versus Moses and Lucy
- Lucy slightly better than Joshua in isolation in out-of-domain, for in-domain Joshua
- Improvement of system was more obvious for out of domain tests
- The hierarchical approach is able to capture the unseen information when RBT system delivers it even when it is only represented vaguely in the translations
- The system not only made better selection of phrase translations provided by Lucy, but also adjusted the translations with more well-formed overall syntactic structures close to the RBMT translation

Chen, Eisele 2010[1]
### Translation Examples

#### In-domain

<table>
<thead>
<tr>
<th>Source</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ich möchte Sie daran erinnern, dass sich unter unseren Verbündeten entschiedene Befürworter dieser Steuer befinden.</td>
<td>Let me remind you that our allies include fervent supporters of this tax.</td>
</tr>
<tr>
<td>Lucy</td>
<td>I would like to remind you of there being decisive proponents of this tax among our allies.</td>
</tr>
<tr>
<td>Moses</td>
<td>I would like to remind you that under our allies are strong supporters of this tax.</td>
</tr>
<tr>
<td>+Lucy</td>
<td>I would like to remind you that there are among our allies in favour of this tax.</td>
</tr>
<tr>
<td>Joshua</td>
<td>I would like to remind you that, under our allies are strong supporters of this tax.</td>
</tr>
<tr>
<td>+Lucy</td>
<td>I would like to remind you that there are strong supporters of this tax among our allies.</td>
</tr>
</tbody>
</table>

#### Out-of-domain

<table>
<thead>
<tr>
<th>Source</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>So kooperieren die Hochschulen schon aus Tradition mit den Nachbarländern.</td>
<td>The university-level institutions’ cooperation with the neighboring countries, for instance, is part of a tradition.</td>
</tr>
<tr>
<td>Lucy</td>
<td>So the colleges co-operate already from tradition with the neighbor countries closely.</td>
</tr>
<tr>
<td>Moses</td>
<td>So the universities from tradition cooperate closely with the neighbouring countries.</td>
</tr>
<tr>
<td>+Lucy</td>
<td>So the colleges co-operate closely with the neighbouring already from tradition.</td>
</tr>
<tr>
<td>Joshua</td>
<td>So cooperate closely with the neighbouring the universities from tradition.</td>
</tr>
<tr>
<td>+Lucy</td>
<td>So the universities, already from tradition, co-operate closely with the neighbouring countries.</td>
</tr>
</tbody>
</table>

**Figure:** Translation examples (Chen, Eisele 2010[^1])
Possibilities for Future Work

- Integration of commercial RBMT system with hierarchical SMT system by extracting SCFG rules from RBMT systems, which inherits lexicons from both sub-systems, showed a 1.5 BLEU score improvement versus other hybrid systems and or the systems in isolation.
- Evidence that this leads in a promising direction.
- A tighter integration of SMT and RBMT will eventually lead to significant progress.

Chen, Eisele 2010[1]
References: Journals and Presentations


References: Web


