The RWTH Machine Translation System for IWSLT 2008

David Vilar, Daniel Stein, Yuqi Zhang, Evgeny Matusov, Arne Mause, Oliver Bender, Saab Mansour and Hermann Ney

IWSLT’08 21. October, 2008

Human Language Technology and Pattern Recognition Lehrstuhl für Informatik 6 Computer Science Department RWTH Aachen University, Germany
1 Introduction

- Combination of phrase-based and hierarchical SMT systems
- Chinese-to-English and Arabic-to-English
- Investigated the effect of
  - different preprocessing techniques
  - reordering methods (including reordering of speech lattices)
  - syntax-based enhancements
- System ranked 6th in CE (all conditions) and 3rd in AE (all conditions)
- Combination of AE and CE outputs
Outline

1 Introduction
2 Translation Models
3 Extensions
4 Experimental Results
5 Conclusions
2 Translation Models

2.1 Phrase-based Model

- Well-known model
- Scores computed by relative frequencies
- Two different reordering models (depending on language pair)
  - IBM Reordering
  - Jump Reordering
2.2 Hierarchical Model

- Extension of the phrase-based model
- Allow for “gaps” in the phrases
- Formalized as a CF grammar (translation as parsing process)
- Example rules:

\[
X \rightarrow \langle 中 \ X^0 \ 那个 \ X^1, \text{It's the } X^1 \text{ in the } X^0 \rangle \\
X \rightarrow \langle 也要 \ X^0 \ 一些 \ X^1, \text{like to } X^0 \text{ some } X^1 \text{ too} \rangle
\]
2.3 Common Models

- Word-based Lexicon Model
- Target Language Model (6-gram, Kneser-Ney discounting)
- Phrase Count Features
- Phrase Penalty
- Word Penalty
3 Extensions

3.1 Syntactical Features

- Extension for the hierarchical model
- Additional (soft) feature extracted at training time
- A rule is “syntactically consistent” if the “involved” original phrases correspond to yields of a syntax tree
- Done for source and target part independently
- Possibility of smoothing the “syntactic constraints”
- Detailed description in [Vilar et al. 2008]
3.2 Chunk-based Reordering for Chinese

- Reordering of the Chinese source sentence
- Syntactic chunk-level rules, automatically learned from the training data
- Reordered possibilities represented as $n$-best lists (with small $n$)
- Each reordering scored with the product of the probability of each of the rules
- Reordered training data added to the original data
- Detailed description in [Zhang et al. 2007]
3.3 Source Preprocessing

Chinese

- Unigram segmenter obtains better results than ictclass
- LDC-like segmentation without text normalization
- Maximize the joint probability of all the words in the sentence
- Splitting long sentence pairs
- Detailed description in [Xu et al. 2008]
Experiments with MADA and MorphTagger (POS-tagging tool)

Three segmentation schemes

- Splitting only the prefixes w+, l+, k+, b+, s+ (PRE)
- Additionally splitting the determiner Al+ (PRE+DET)
- Additionally splitting the pronominal suffixes (PREF+SUF)

Tested normalizing Yaa and Alef

Best results:

- MorphTagger: PRE+SUF and no normalization
- MADA: PRE and normalization
3.4 Translation of Speech Lattices

- Translation of word lattices including reordering
- Acoustic and source language model scores
- Cardinality synchronous search
  - Define cardinality in terms of “slots” (CN-like)
  - Allow for reordering without the over-generalization of CN
- Mapping from ASR vocabulary to MT vocabulary (segmentation)
- No improvements on this task (regretfully)
- Detailed description in [Matusov et al. 2008]
3.5 System Combination

- Approach used in last year’s evaluation
- Build a confusion network for each sentence
  - Select one system as primary system
  - Align the single-best output of this system with the other hypotheses
  - Build a confusion network
  - Repeat with each system as primary
- The resulting confusion networks are joined into a word graph
- Weight with system specific factors and a trigram LM trained on the MT hypotheses
- Detailed description in [Matusov et al. 2006]
4 Experimental Results

- Arabic-to-English and Chinese-to-English translation directions
- Provided training data + HIT Corpus for CE
  - Selected sentences with 60% of the words in the IWSLT data
- Preprocessing of English:
  - Tokenization
  - Expansion of contractions
- GIZA++ for alignments
  - Tested different variants of word classes, model sequences and combination heuristics
- Optimized for BLEU on IWSLT 2004 eval data
- IWSLT 2005 eval data for system combination
## 4.1 Chinese-to-English

### BTEC Task

<table>
<thead>
<tr>
<th>System</th>
<th>BLEU</th>
<th>TER</th>
<th>WER</th>
<th>PER</th>
</tr>
</thead>
<tbody>
<tr>
<td>System Combination</td>
<td>46.1</td>
<td>37.7</td>
<td>43.9</td>
<td>39.4</td>
</tr>
<tr>
<td>Phrase Based (PBT)</td>
<td>42.5</td>
<td>36.6</td>
<td>45.3</td>
<td>40.6</td>
</tr>
<tr>
<td>PBT + Chunk Reordering</td>
<td>42.6</td>
<td>39.9</td>
<td>47.8</td>
<td>42.4</td>
</tr>
<tr>
<td>PBT + New Segmentation</td>
<td>44.3</td>
<td>40.3</td>
<td>47.3</td>
<td>42.0</td>
</tr>
<tr>
<td>Hierarchical</td>
<td>41.2</td>
<td>41.5</td>
<td>48.1</td>
<td>42.7</td>
</tr>
<tr>
<td>Hierarchical + Syntax</td>
<td>41.4</td>
<td>40.6</td>
<td>47.3</td>
<td>42.8</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>System</th>
<th>BLEU</th>
<th>TER</th>
<th>WER</th>
<th>PER</th>
</tr>
</thead>
<tbody>
<tr>
<td>System Combination</td>
<td>39.7</td>
<td>42.5</td>
<td>49.6</td>
<td>44.5</td>
</tr>
<tr>
<td>Phrase Based (PBT)</td>
<td>37.3</td>
<td>41.2</td>
<td>50.0</td>
<td>45.1</td>
</tr>
<tr>
<td>PBT + Chunk Reordering</td>
<td>38.5</td>
<td>42.8</td>
<td>51.2</td>
<td>46.4</td>
</tr>
<tr>
<td>Hierarchical</td>
<td>31.6</td>
<td>49.6</td>
<td>56.5</td>
<td>49.5</td>
</tr>
<tr>
<td>Hierarchical + Syntax</td>
<td>36.6</td>
<td>44.1</td>
<td>51.4</td>
<td>47.0</td>
</tr>
<tr>
<td>Lattices</td>
<td>32.2</td>
<td>48.6</td>
<td>57.1</td>
<td>51.5</td>
</tr>
</tbody>
</table>
## Challenge Task

### CRR

<table>
<thead>
<tr>
<th>System</th>
<th>BLEU</th>
<th>TER</th>
<th>WER</th>
<th>PER</th>
</tr>
</thead>
<tbody>
<tr>
<td>System Combination</td>
<td>39.1</td>
<td>40.7</td>
<td>48.3</td>
<td>44.1</td>
</tr>
<tr>
<td>Phrase Based (PBT)</td>
<td>32.1</td>
<td>42.7</td>
<td>51.9</td>
<td>47.8</td>
</tr>
<tr>
<td>PBT + Chunk Reordering</td>
<td>32.6</td>
<td>43.6</td>
<td>52.5</td>
<td>48.5</td>
</tr>
<tr>
<td>PBT + New Segmentation</td>
<td>37.2</td>
<td>41.8</td>
<td>49.3</td>
<td>44.5</td>
</tr>
<tr>
<td>Hierarchical</td>
<td>30.7</td>
<td>47.1</td>
<td>54.6</td>
<td>48.9</td>
</tr>
<tr>
<td>Hierarchical + Syntax</td>
<td>30.2</td>
<td>45.5</td>
<td>53.6</td>
<td>48.5</td>
</tr>
</tbody>
</table>

### ASR

<table>
<thead>
<tr>
<th>System</th>
<th>BLEU</th>
<th>TER</th>
<th>WER</th>
<th>PER</th>
</tr>
</thead>
<tbody>
<tr>
<td>System Combination</td>
<td>34.3</td>
<td>43.6</td>
<td>51.1</td>
<td>46.1</td>
</tr>
<tr>
<td>Phrase Based (PBT)</td>
<td>27.8</td>
<td>46.0</td>
<td>55.4</td>
<td>51.1</td>
</tr>
<tr>
<td>PBT + Chunk Reordering</td>
<td>29.4</td>
<td>45.7</td>
<td>55.0</td>
<td>50.5</td>
</tr>
<tr>
<td>Hierarchical</td>
<td>26.4</td>
<td>51.0</td>
<td>59.2</td>
<td>51.9</td>
</tr>
<tr>
<td>Hierarchical + Syntax</td>
<td>30.2</td>
<td>45.6</td>
<td>53.7</td>
<td>48.6</td>
</tr>
<tr>
<td>Lattices</td>
<td>25.0</td>
<td>56.6</td>
<td>62.8</td>
<td>56.7</td>
</tr>
</tbody>
</table>
4.2 Arabic-to-English

<table>
<thead>
<tr>
<th>System</th>
<th>BLEU</th>
<th>TER</th>
<th>WER</th>
<th>PER</th>
</tr>
</thead>
<tbody>
<tr>
<td>System Combination</td>
<td>53.5</td>
<td>33.0</td>
<td>37.6</td>
<td>33.9</td>
</tr>
<tr>
<td>PBT + MADA</td>
<td>50.0</td>
<td>33.7</td>
<td>39.7</td>
<td>36.0</td>
</tr>
<tr>
<td>PBT + MorphTagger</td>
<td>51.8</td>
<td>33.8</td>
<td>38.1</td>
<td>33.9</td>
</tr>
<tr>
<td>Hierarchical + MADA</td>
<td>49.2</td>
<td>36.6</td>
<td>41.3</td>
<td>36.7</td>
</tr>
<tr>
<td>Hierarchical + MorphTagger</td>
<td>49.3</td>
<td>35.9</td>
<td>41.3</td>
<td>38.0</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>System</th>
<th>BLEU</th>
<th>TER</th>
<th>WER</th>
<th>PER</th>
</tr>
</thead>
<tbody>
<tr>
<td>System Combination</td>
<td>44.5</td>
<td>37.6</td>
<td>43.4</td>
<td>39.9</td>
</tr>
<tr>
<td>PBT + MADA</td>
<td>42.6</td>
<td>38.2</td>
<td>45.3</td>
<td>41.7</td>
</tr>
<tr>
<td>PBT + MorphTagger</td>
<td>44.0</td>
<td>38.0</td>
<td>43.4</td>
<td>39.4</td>
</tr>
<tr>
<td>Hierarchical + MADA</td>
<td>41.3</td>
<td>42.1</td>
<td>47.7</td>
<td>42.7</td>
</tr>
<tr>
<td>Hierarchical + MorphTagger</td>
<td>41.3</td>
<td>40.7</td>
<td>47.2</td>
<td>43.9</td>
</tr>
</tbody>
</table>

Note: bug in the hierarchical system (Corrected score for CRR: 54.1%)
4.3 Arabic&Chinese-to-English

- System combination of the best performing systems for both language pairs

<table>
<thead>
<tr>
<th>CRR</th>
<th>BLEU</th>
<th>TER</th>
<th>WER</th>
<th>PER</th>
</tr>
</thead>
<tbody>
<tr>
<td>System Combination</td>
<td>56.2</td>
<td>31.7</td>
<td>36.0</td>
<td>32.6</td>
</tr>
</tbody>
</table>
5 Conclusions

- Presented RWTH system for the IWSLT 2008 evaluation
- Combination of different statistical machine translation approaches
  - Phrase-based and hierarchical systems + extensions
- Combination of Arabic-to-English and Chinese-to-English systems increases performance