Statistical Machine Transliteration with Multi-to-Multi Joint Source Channel Model

Rui Wang
with Yu Chen and Yi Zhang
DFKI GmbH, Germany
Outline

• Motivation
• System Overview
• Evaluation Results
• Conclusion
Motivation

• Previous Work
  – Joint Source-Channel
    – [Li et al., 2004], [Li et al., 2007]
  – Pinyin
  – Statistical Machine Translation (SMT)

• Our Setup
  – Multi-to-Multi Joint Source-Channel
  – System integration
Transliteration in SMT

- **EuroMatrixPlus (EU FP7)**
  - English, German, French, Spanish ↔ Chinese
  - Hybrid MT

- **Named-Entity translation**
  - Statistical
  - Rule-based
  - Hybrid
System Overview
Phrase-based SMT

- Translation model
- Language model

\[
\arg\max_{t} P(t \mid s) = \arg\max_{t} (P(t)P(s \mid t))
\]
Joint Source-Channel Model

- [Li et al., 2004]

\[
P(s,t,\alpha) = \sum_{k=1}^{K} P(<e,c>_{k} | <e,c>_{k-n+1})
\]

\[
\bar{t} = \arg \max_{s,\alpha} P(s,t,\alpha) \\
\bar{s} = \arg \max_{t,\alpha} P(s,t,\alpha)
\]
Multi-to-Multi JSC

<table>
<thead>
<tr>
<th>English</th>
<th>Chinese</th>
</tr>
</thead>
<tbody>
<tr>
<td>A/JA/X</td>
<td>埃/甲/克斯</td>
</tr>
<tr>
<td>A/BA/STE/NIA</td>
<td>阿/巴/斯蒂/尼亚</td>
</tr>
<tr>
<td>AHL/BERG</td>
<td>阿尔/伯格</td>
</tr>
</tbody>
</table>

- N-gram → 1-gram
- Random start → Normalize all
Experiments

• Preprocessing

• SMT settings
  ▫ Parameter tuning

• Results
Preprocessing

• Tokenization

• No-phonetic mapping

• Separate training sets
SMT Setting

• Models
  ▫ Translation model
  ▫ “Language model” SRILM [Stolcke, 2002]
  ▫ Length penalty

• Moses Decoder
  ▫ [Koehn et al., 2007]
  ▫ No distortion
Parameter Tuning

- **MERT**
  - [Och, 2003]
  - Z-MERT [Zaidan, 2009]

- **Development Set**
  - Randomly 500

- **F-score-based tuning**
  - Instead of BLEU
(Official) Results

<table>
<thead>
<tr>
<th>Tasks</th>
<th>System</th>
<th>ACC</th>
<th>Mean F</th>
<th>MRR</th>
<th>Map_ref</th>
</tr>
</thead>
<tbody>
<tr>
<td>English-to-Chinese</td>
<td>M2MJSC +PBSMT</td>
<td>0.320</td>
<td>0.674</td>
<td>0.397</td>
<td>0.308</td>
</tr>
<tr>
<td>English-to-Chinese</td>
<td>M2MJSC</td>
<td>0.260</td>
<td>0.638</td>
<td>0.340</td>
<td>0.251</td>
</tr>
<tr>
<td>Chinese-to-English</td>
<td>M2MJSC +PBSMT</td>
<td>0.133</td>
<td>0.746</td>
<td>0.210</td>
<td>0.133</td>
</tr>
<tr>
<td>Chinese-to-English</td>
<td>M2MJSC</td>
<td>0.117</td>
<td>0.731</td>
<td>0.177</td>
<td>0.117</td>
</tr>
</tbody>
</table>

Chiang Mai, Thailand 11/12/11
# Additional Results

<table>
<thead>
<tr>
<th>Tasks</th>
<th>LM</th>
<th>ACC</th>
<th>Mean F</th>
<th>MRR</th>
<th>Map_ref</th>
</tr>
</thead>
<tbody>
<tr>
<td>English-to-Chinese</td>
<td>3-gram</td>
<td>30.62</td>
<td>67.30</td>
<td>38.63</td>
<td>29.71</td>
</tr>
<tr>
<td>English-to-Chinese</td>
<td>5-gram</td>
<td>31.71</td>
<td>67.71</td>
<td>39.43</td>
<td>30.70</td>
</tr>
<tr>
<td>English-to-Chinese</td>
<td>7-gram</td>
<td>31.62</td>
<td>67.61</td>
<td>39.46</td>
<td>30.66</td>
</tr>
<tr>
<td>Chinese-to-English</td>
<td>3-gram</td>
<td>11.67</td>
<td>73.06</td>
<td>18.29</td>
<td>11.68</td>
</tr>
<tr>
<td>Chinese-to-English</td>
<td>5-gram</td>
<td>12.48</td>
<td>73.29</td>
<td>18.82</td>
<td>12.47</td>
</tr>
<tr>
<td>Chinese-to-English</td>
<td>7-gram</td>
<td>12.97</td>
<td>73.89</td>
<td>19.48</td>
<td>12.97</td>
</tr>
</tbody>
</table>
Conclusion

- First participation
- M2M-JSC
- System integration (20%)
- LM effectiveness
Future Work

- Richer feature set for M2MJSC
- Different models for SMT
- Additional knowledge
  - Pinyin
  - Dictionaries
- System combination
  - [Chen et al., 2009]
Xie Xie!

Questions?