Motivation

- ASR output does not contain punctuation marks
- MT systems are trained on text data with punctuation
- prediction errors affect translation quality
- loss of up to 4 BLEU points if punctuation marks need to be predicted, compared to correct punctuation in the input

<table>
<thead>
<tr>
<th>system</th>
<th>MT dev</th>
<th>MT test</th>
</tr>
</thead>
<tbody>
<tr>
<td>+s2t TED trip.</td>
<td>27.5/57.0</td>
<td>30.8/50.9</td>
</tr>
<tr>
<td>correct punctuation</td>
<td>24.0/61.7</td>
<td>26.6/55.9</td>
</tr>
</tbody>
</table>

Comparison of the Translation Quality

- IWslt 2011 English-to-French speech translation of talks [Federico et al., IWslt 2011]
- uses +s2t TED trip. from English-French MT for all punctuation prediction strategies [Wuebker et al., IWslt 2011]
- system combination [Matusov et al, EACL 2006]
- combine translation output from multiple punctuation prediction schemes

<table>
<thead>
<tr>
<th>system</th>
<th>SLT dev</th>
<th>SLT test</th>
</tr>
</thead>
<tbody>
<tr>
<td>IMPPLICIT</td>
<td>18.0</td>
<td>69.5</td>
</tr>
<tr>
<td>FULLPUNCT (H-NGRAM)</td>
<td>18.2</td>
<td>69.3</td>
</tr>
<tr>
<td>FULLPUNCT (PPMT)</td>
<td>18.3</td>
<td>69.2</td>
</tr>
<tr>
<td>NOPUNCT (H-NGRAM)</td>
<td>17.3</td>
<td>67.9</td>
</tr>
<tr>
<td>NOPUNCT (PPMT)</td>
<td>17.8</td>
<td>69.0</td>
</tr>
<tr>
<td>system comb.</td>
<td>18.5</td>
<td>68.3</td>
</tr>
</tbody>
</table>

Introduction

- in this work, we consider all kinds of punctuation
- sentence-end punctuation marks
- commas
- parentheses and quotation marks
- punctuation prediction is performed via
  - tool from the SRI LM toolkit [Stolcke, ICSLP 2002]
  - statistical machine translation [Hassan et al., IWslt 2007]
  - comparison and combination of different methods
  - applied in the IWslt 2011 evaluation campaign

Comparison of the Punctuation Prediction Accuracy

- three different stages at which prediction is done
  - before translation in the source language (FULLPUNCT)
    - no modification to the training data or the translation system
    - prediction errors can affect the translation
  - during translation implicitly (IMPLICIT)
    - removing all punctuation marks from the source language data
    - re-extracting phrase and word lexicon models
    - prediction and translation are not separate
  - after translation in the target language (NOPUNCT)
    - all punctuation marks are removed from the training data as well as from the development and test sets
    - translation model and target language model have to be rebuilt
    - translation produces errors, make the punctuation prediction less accurate

<table>
<thead>
<tr>
<th>tool</th>
<th>class 1</th>
<th>class 1.1</th>
<th>class 1.2</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Prec.</td>
<td>Rec.</td>
<td>F1</td>
</tr>
<tr>
<td>H-NGRAM</td>
<td>87.9</td>
<td>85.0</td>
<td>88.4</td>
</tr>
<tr>
<td>PPMT</td>
<td>88.2</td>
<td>81.7</td>
<td>84.8</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>tool</td>
<td>class 2</td>
<td>class 3</td>
<td>all punct.</td>
</tr>
<tr>
<td></td>
<td>Prec.</td>
<td>Rec.</td>
<td>F1</td>
</tr>
<tr>
<td>H-NGRAM</td>
<td>83.5</td>
<td>44.8</td>
<td>58.3</td>
</tr>
<tr>
<td>PPMT</td>
<td>80.6</td>
<td>59.3</td>
<td>68.3</td>
</tr>
</tbody>
</table>

Example

- with hidden-NGRAM tool from the SRI LM toolkit (H-NGRAM)
  - standard setting with 9-gram language model
- with statistical machine translation (PPMT)
  - based on phrase-based MT system
  - additional features besides the language model
  - translate from unpunctuated to punctuated text
  - system is tuned with standard MERT on BLEU

- with hidden-NGRAM tool from the SRI LM toolkit (H-NGRAM)
  - standard setting with 9-gram language model
- with statistical machine translation (PPMT)
  - based on phrase-based MT system
  - additional features besides the language model
  - translate from unpunctuated to punctuated text
  - system is tuned with standard MERT on BLEU

Punctuation Prediction

- with hidden-NGRAM tool from the SRI LM toolkit (H-NGRAM)
  - standard setting with 9-gram language model
- with statistical machine translation (PPMT)
  - based on phrase-based MT system
  - additional features besides the language model
  - translate from unpunctuated to punctuated text
  - system is tuned with standard MERT on BLEU

<table>
<thead>
<tr>
<th>system</th>
<th>tool</th>
<th>example</th>
</tr>
</thead>
<tbody>
<tr>
<td>pseudo ASR output</td>
<td>-</td>
<td>they say &quot;The plants talk to us&quot;</td>
</tr>
<tr>
<td>reference</td>
<td>-</td>
<td>they say &quot;The plants talk to us.&quot;</td>
</tr>
<tr>
<td>FULLPUNCT</td>
<td>H-NGRAM</td>
<td>they say &quot;The plants talk to us&quot;</td>
</tr>
<tr>
<td>FULLPUNCT</td>
<td>PPMT</td>
<td>they say &quot;The plants talk to us.&quot;</td>
</tr>
</tbody>
</table>

Conclusion

- compared different approaches for predicting punctuation in a speech translation setting
  - PPMT outperformed H-NGRAM
  - FULLPUNCT (PPMT) slightly better than the implicit method
  - main advantage of FULLPUNCT: no modification to the translation system
  - system combination improved translation quality further
- future work:
  - investigate special features for parentheses or quotes
  - try different optimization criteria, e.g. F-measure or WER

Lehrstuhl für Informatik 6 – Computer Science Department – RWTH Aachen University – Aachen, Germany

e-mail: <surname>@cs.rwth-aachen.de

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