Domain Adaptation For Subtitles You Can Take Seriously

Muhammad Ali Tahir

transLectures Workshop, Ljubljana
April 22nd, 2014
Outline

1. Task definition for transLectures
2. Main challenges and opportunities
3. Adaptation of acoustic models
4. Adaptation of language models
5. Adaptation of translation models
6. Experimental results
7. Tools for adaptation
8. Conclusions
Task Definition

- Automatic transcription and translation of video lectures

- Basic components of the system
  - Audio $\rightarrow$ Text (automatic speech recognition)
  - Text $\rightarrow$ Text in another language (statistical machine translation)

- Can we make use of the extra information?
  - Slides $\rightarrow$ Slides text (helpful for speech recognition)
Translectures: Challenges and Opportunities

▶ Lecture recognition and translation:
  ▶ Heterogeneous conditions (non-native speakers, echo in lecture rooms ...)
  ▶ Broad range of topics (technical terms ...)
  ▶ Need for adaptation methods

▶ Large amounts of data for transcription and translation:
  ▶ videolectures.net and poliMedia
  ▶ Recognition of 7,000 hours of speech
  ▶ Translation into multiple languages, e.g., from English into {French, German, Slovenian, Spanish}
  ▶ Focus on efficient adaptation methods

▶ Availability of lecture slides
  ▶ Exploitation of prior knowledge
Automatic Speech Recognition

▶ Conversion of audio into text
▶ Four main components
  ▶ Signal analysis
  ▶ Acoustic model
  ▶ Language model
  ▶ Search

Speech Input

Feature Extraction

Feature Vectors $x_1...x_T$

Global Search Process:

$$\text{maximize} \quad p(w_1...w_N) \cdot p(x_1...x_T | w_1...w_N)$$

over $w_1...w_N$

Recognized Word Sequence

$\{w_1...w_N\}_{\text{opt}}$

Acoustic Model
- subword units
- pronunciation lexicon

Language Model

$\ p(w_1...w_N)$
Statistical Machine Translation

- Conversion of text in one language into another language
- Four main components
  - Translation model (lexicon + alignment)
  - Language model
  - Transformations (pre-/post-processing)
  - Search

Global Search Process:

\[
\maximize \quad p(e'_{1}) \cdot p(r'_{1}|e'_{1})
\]

over \( e'_{1} \)

Source Language Text

Transformation

\( r'_{1} \)

Global Search Process:

\[
p(r'_{1}|e'_{1})
\]

\[
p(e'_{1})
\]

Target Language Text

Lexicon Model

Alignment Model

Language Model

Tahir: Domain Adaptation 6 / 24 April 22nd, 2014
Performance Metrics

► Automatic speech recognition
  ▶ Word error rate (WER)

\[
\text{deletions} + \text{insertions} + \text{substitutions} \\
\text{reference words}
\]

<table>
<thead>
<tr>
<th></th>
<th>E</th>
<th>L</th>
<th>E</th>
<th>P</th>
<th>H</th>
<th>A</th>
<th>N</th>
<th>T</th>
</tr>
</thead>
<tbody>
<tr>
<td>E</td>
<td>0</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>5</td>
<td>6</td>
<td>7</td>
</tr>
<tr>
<td>L</td>
<td>6</td>
<td>7</td>
<td>8</td>
<td>9</td>
<td>10</td>
<td>11</td>
<td>12</td>
<td>13</td>
</tr>
<tr>
<td>P</td>
<td>3</td>
<td>4</td>
<td>5</td>
<td>6</td>
<td>7</td>
<td>8</td>
<td>9</td>
<td>10</td>
</tr>
<tr>
<td>H</td>
<td>11</td>
<td>12</td>
<td>13</td>
<td>14</td>
<td>15</td>
<td>16</td>
<td>17</td>
<td>18</td>
</tr>
<tr>
<td>A</td>
<td>19</td>
<td>20</td>
<td>21</td>
<td>22</td>
<td>23</td>
<td>24</td>
<td>25</td>
<td>26</td>
</tr>
<tr>
<td>N</td>
<td>27</td>
<td>28</td>
<td>29</td>
<td>30</td>
<td>31</td>
<td>32</td>
<td>33</td>
<td>34</td>
</tr>
<tr>
<td>T</td>
<td>35</td>
<td>36</td>
<td>37</td>
<td>38</td>
<td>39</td>
<td>40</td>
<td>41</td>
<td>42</td>
</tr>
</tbody>
</table>

► Statistical machine translation
  ▶ Bilingual evaluation understudy (BLEU)
Detailed Project Goals

▶ Partners involved:
  ▶ RWTH Aachen University (RWTH)
  ▶ European Media Laboratory GmbH (EML)
  ▶ Polytechnic University of Valencia (UPVLC)
  ▶ Xerox Research Center Europe (XEROX)

▶ Massive adaptation of models in ASR and SMT
  ▶ Adaptation of acoustic models
  ▶ Adaptation of language models
  ▶ Adaptation of translation models
Investigations on acoustic model adaptation methods

- CMLLR (constrained maximum likelihood linear regression)
- MLLR (maximum likelihood linear regression)
- MAP (maximum a-posteriori adaptation)

Adaptation by neural networks

Directions of research:

- Shallow vs. deep
- Tandem vs. hybrid
- Mono- vs. multilingual
Adaptation of Acoustic Models

Speech Input

Acoustic Analysis

CMLLR

MLLR

Global Search:
maximize
Pr (x₁...xₜ | w₁...w₆)

Pr (w₁...w₆) • Pr (x₁...xₜ | w₁...w₆)

over w₁...w₆

Recognized Word Sequence

Acoustic Model:
- subword units
- pronunciation lexicon

Language Model
Multilingual Neural Network Adaptation

- Adaptation with data from other languages
- Dimensionality reduction by bottleneck layer
## Languages and Training Data

<table>
<thead>
<tr>
<th>Repository</th>
<th>Language</th>
<th>AM Hours</th>
<th>LM Words</th>
</tr>
</thead>
<tbody>
<tr>
<td>videolectures.net</td>
<td>English</td>
<td>1006</td>
<td>6600 M</td>
</tr>
<tr>
<td></td>
<td>Slovenian</td>
<td>90</td>
<td>75 M</td>
</tr>
<tr>
<td>poliMedia</td>
<td>Spanish</td>
<td>390</td>
<td>1600 M</td>
</tr>
<tr>
<td></td>
<td>Catalan</td>
<td>49</td>
<td>141 M</td>
</tr>
</tbody>
</table>
Results for Acoustic Model Adaptation

▶ Improvements by acoustic model adaptation only

<table>
<thead>
<tr>
<th>Language</th>
<th>without AM adaptation</th>
<th>with AM adaptation</th>
</tr>
</thead>
<tbody>
<tr>
<td>English</td>
<td>27.9 %</td>
<td>22.8 %</td>
</tr>
<tr>
<td>Spanish</td>
<td>17.0 %</td>
<td>15.6 %</td>
</tr>
<tr>
<td>Slovenian</td>
<td>45.4 %</td>
<td>38.1 %</td>
</tr>
</tbody>
</table>

▶ Additional gains by system combination (Spanish)

▷ System 1 WER: 17.0%
▷ System 2 WER: 15.6%
▷ System 1 + System 2 WER: 15.3%
Adaptation of Language Models

► Adaptation by interpolation with in-domain data

► Refinement: adaptation with slides
  ▶ Adapt LM with all available slides
  ▶ Adapt LM with specific slides of the given lecture
  ▶ Improved OCR (with help from other project)

► Language model cache
Results for Language Model Adaptation

Word error rate results for language model adaptation

<table>
<thead>
<tr>
<th>Language</th>
<th>AM Adaptation</th>
<th>AM+LM Adaptation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Catalan</td>
<td>–</td>
<td>35.5%</td>
</tr>
<tr>
<td>English</td>
<td>22.8%</td>
<td>21.2%</td>
</tr>
<tr>
<td>Spanish</td>
<td>15.6%</td>
<td>13.5%</td>
</tr>
<tr>
<td>Slovenian</td>
<td>38.1%</td>
<td>36.2%</td>
</tr>
</tbody>
</table>

Improvements on top of acoustic adaptation
Impact of More Data vs. Adaptation

- How much do additional data help?
- Case study for English ASR, word error rate results

<table>
<thead>
<tr>
<th>System</th>
<th>WER</th>
</tr>
</thead>
<tbody>
<tr>
<td>Baseline</td>
<td>27.9 %</td>
</tr>
<tr>
<td>+AM Adaptation</td>
<td>23.1 %</td>
</tr>
<tr>
<td>+Additional Data</td>
<td>22.8 %</td>
</tr>
<tr>
<td>+LM Adaptation</td>
<td>21.2 %</td>
</tr>
</tbody>
</table>
Improvements for Automatic Speech Recognition

WER (ASR)

- EnU
- Sl
- Ca
- EnR
- EsU
- EsR

M12 M18 M24
Adaptation of Translation Models

Efforts on LM adaptation methods:
- Adaptation with more in-domain data
- Domain specific LM array
- Lexical coverage features

Data weighting and data selection

Measure similarity to target domain
- Using the language model
- Using the translation model

Data selection based on text that is being translated
## Language Pairs and Training Data

<table>
<thead>
<tr>
<th>Repository</th>
<th>Task</th>
<th>Sentences</th>
</tr>
</thead>
<tbody>
<tr>
<td>poliMedia</td>
<td>Es→En</td>
<td>17.9 M</td>
</tr>
<tr>
<td></td>
<td>En→Es</td>
<td>17.9 M</td>
</tr>
<tr>
<td></td>
<td>En→Fr</td>
<td>38.7 M</td>
</tr>
<tr>
<td>videolectures.net</td>
<td>En→De</td>
<td>12.8 M</td>
</tr>
<tr>
<td></td>
<td>En→Sl</td>
<td>13.4 M</td>
</tr>
<tr>
<td></td>
<td>Sl→En</td>
<td>13.4 M</td>
</tr>
</tbody>
</table>
# Results for Translation Model Adaptation

## Translation quality in BLEU

<table>
<thead>
<tr>
<th>Task</th>
<th>Baseline</th>
<th>TM Adaptation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Es→En</td>
<td>26.0</td>
<td>27.3</td>
</tr>
<tr>
<td>En→Es</td>
<td>33.5</td>
<td>35.4</td>
</tr>
<tr>
<td>En→Fr</td>
<td>32.4</td>
<td>35.0</td>
</tr>
<tr>
<td>En→De</td>
<td>19.3</td>
<td>23.0</td>
</tr>
<tr>
<td>En→Sl</td>
<td>12.0</td>
<td>15.9</td>
</tr>
<tr>
<td>Sl→En</td>
<td>15.7</td>
<td>21.8</td>
</tr>
</tbody>
</table>
Improvements for Statistical Machine Translation

![Diagram showing BLEU scores for different language pairs over time]

- **EnEs**: Black line with circles
- **EnFrX**: Red line with circles
- **EsEn**: Black line with circles
- **EnDe**: Blue line with circles
- **SlEn**: Black line with circles
- **EnSl**: Black line with circles

**X-axis**: M12, M18, M24

**Y-axis**: BLEU (MT)
Tools for Adaptation

► EML
  ▶ ASR: EML Transcription Platform

► RWTH
  ▶ ASR: RASR
  ▶ SMT: Jane

► UPLVC
  ▶ ASR: TLK

► XEROX
  ▶ SMT: TunaTon Toolkit
Conclusions

► ASR
  ▶ Large improvements of 20-30 % relative for all languages
  ▶ English and Spanish: accurate enough transcriptions
  ▶ AM adaptation improved by neural networks
  ▶ LM adaptation improved by slides (better OCR)
  ▶ Initial release of UPVLC TLK toolkit

► SMT
  ▶ Either high quality scores or large improvements
  ▶ Translation model adaptation improved by data selection and data weighting

► Future Work
  ▶ Improved Slovenian ASR, Slovenian and German SMT
Thank you for your attention!
Ongoing Work

WER (ASR)

- M06
- M12
- M18
- M24
- M30

BLEU (MT)

- EnU
- Sl
- Ca
- EnR
- EsU
- EsR
- EnEs
- EnFrX
- EnDe
- SlEn
- EnSl

Tahir: Domain Adaptation 25 / 24 April 22nd, 2014