Immediate Adaptation to User Corrections in Post-Editing SMT

Patrick Simianer, Sariya Karimova, Stefan Riezler
Heidelberg University, Germany

Oct 28, 2016
iMT 2016 : AMTA 2016 Workshop on Interacting with Machine Translation
1 Motivation

2 Proposed approach

3 User study
1 Motivation

2 Proposed approach

3 User study
Motivation

User-adaptation in computer-aided translation (CAT) is crucial

1. To overcome domain shifts between training data and translated materials
2. To prevent frustrations with translation technology, e.g. related to post-editing
3. To boost efficiency and (possibly) quality
[title] Sheathed element glow plug

[abstract segment #1] A sheathed element glow plug (1) is to be placed inside a chamber (3) of an internal combustion engine.

[abstract segment #2] The sheathed element glow plug (1) comprises a heating body (2) that has a glow tube (6) connected to a housing (4).
A sheathed element glow plug (1) serves for arrangement in a chamber of an internal combustion engine.

The sheathed element glow plug comprises a heating body (2) which has a glow tube (5) and a heating coil (8) which is arranged in the glow tube (5).
Motivation

• Translation memories naturally adapt to their users, this raises expectations
  → But updating SMT-based CAT systems is not straight-forward

• Adaptation by re-training (overnight) is useful
  → But it can’t help during translation sessions, as it’s a slow process

• Online adaptive SMT is well studied and there are even products\(^1\) that implement it
  → But most research is theoretical, user studies are scarce
  → Adaptation is potentially imprecise due to automatic alignment methods

\(^1\)lilt.com, SDL Trados
Proposed approach

We present an approach to **online** user adaptive post-editing with precise, **immediate** adaptation:

⇒ By leveraging **user-generated alignments** for phrase-table adaptation

⇒ We evaluate our approach to adaptation in a **user study**
Definition of online adaptation

For each example $t = 1, \ldots, |\mathcal{d}|$

1. Receive input sentence $x_t$
2. Output translation $\hat{y}_t$ from current model
3. Receive user output $y_t$
4. Refine models on $(x, \hat{y}, y)_t$

Figure: Online learning procedure in computer-aided translation
Related work

W/o user study: Bertoldi et al. [2014]², Ortiz-Martínez et al. [2010]³, Wuebker et al. [2015b]⁴

W/ user study: Green et al. [2014]⁵, Denkowski [2015]⁶

Automatic alignment model: Bertoldi et al. [2014], Denkowski [2015], Ortiz-Martínez et al. [2010]

Tuning only: Green et al. [2014], Wuebker et al. [2015b]

---

²Online adaptation to post-edits for phrase-based statistical machine translation
³Online Learning for Interactive Statistical Machine Translation
⁴Hierarchical Incremental Adaptation for Statistical Machine Translation
⁵Human Effort and Machine Learnability in Computer Aided Translation
⁶Machine Translation for Human Translators
Related work – Evaluation

Quality
- Measure BLEU/TER of post-edits wrt. given reference translations (not necessarily meaningful)

Simulated quality
- Measure BLEU/TER of unaltered MT outputs wrt. given reference translations (identical to standard MT evaluation)

Manual effort
1. Measure BLEU/TER of MT outputs wrt. post-edits [HTER]
2. Measure and normalize counts of clicks and keystrokes

Simulated manual effort
1. Measure TER/BLEU wrt. offline created post-edits
2. Use a model of user behavior to estimate number of clicks/keystrokes needed to produce reference translation from MT output
Related work – Evaluation

• Ortiz-Martínez et al. [2010]: Improved simulated quality and simulated manual effort compared to static systems
• Bertoldi et al. [2014]: Improved simulated quality compared to static systems
• Green et al. [2014]: Improved simulated manual effort compared to non-adapted system
• Wuebker et al. [2015b]: Improved simulated quality compared to baseline system
• Denkowski [2015]: Improved simulated quality and manual effort compared to static systems
1 Motivation

2 Proposed approach

3 User study
Example – MT output #1

Sheathed element glow plug

Glühstiftkerze
Example – User correction #1

- sheathed element glow plug $\rightarrow$ Glühkerze
Example – MT output #2

A sheathed element glow plug is to be placed in a chamber of a combustion engine.
Example – User correction #2

Immediately learned translation rules:

- $a_0 \rightarrow \text{eine}$
- is to be placed$_{2,3}$ $X_1 \rightarrow \text{wird } X_1 \text{ eingebaut}$
- a chamber$_5 \rightarrow \text{eine Kammer}$
- of $a_{6,7} \rightarrow \text{eines}$
- combustion engine$_8 \rightarrow \text{Verbrennungsmotors}$
Derived translation rules:

- **in a chamber** $\rightarrow$ in eine Kammer
- **of a combustion engine** $\rightarrow$ eines Verbrennungsmotors
- **in a chamber of a combustion engine** $\Rightarrow$ in eine Kammer eines Verbrennungsmotors
- **in a chamber of $X_1$ combustion engine** $\rightarrow$ in eine Kammer $X_1$ Verbrennungsmotors

...
Example – MT output #3

Spark plug

Zündkerze
Example – MT output #4

Eine Zündkerze wird in den eingebaut Zylinderkopf eines Verbrennungsmotors.
Eine Zündkerze wird in den Zylinderkopf eines Verbrennungsmotors eingebaut.
Approach – Weight updates

- Pairwise ranking updates to weigh many sparse features, e.g. rule ids
- Per coordinate learning rates used to prevent too harsh changes
- Default learning rate for id features of newly extracted rules is the overall median
- Leave-one-out: Derived translation rules are only added to subsequent grammars to prevent overfitting
Approach – Summary

1. (User correction received)
2. Extract immediate corrections from post-edit and alignment and add to current grammar
3. Re-translate input with new grammar to generate k-best list
4. Pairwise ranking update using k-best
5. Add N-grams of post-edit to adaptive language model (following Denkowski et al. [2014])
6. Derive all possible rules from user correction
   
   : 
   
   : 

...
1 Motivation

2 Proposed approach

3 User study
User study – Setup

Subjects
19 students, 13 prospective translators, 6 CS students, 4 different mother tongues

Data
Titles and abstracts of patent documents, filtered by length, clustered by similarity

Environment
Controlled environment in a computer pool, 90 minute sessions

Machine translation
Hierarchical phrase-based system built from title/abstract training data, good baseline translation results

Task
Post-edit about 500 words from English into German, each task is shared by two subjects
## User study – Results

<table>
<thead>
<tr>
<th>response variable</th>
<th>estimated $\Delta$</th>
</tr>
</thead>
<tbody>
<tr>
<td>HBLEU$_{+1}$</td>
<td>+6.8 ± 2.0 [%]</td>
</tr>
<tr>
<td>HTER</td>
<td>−5.3 ± 1.9 [%]</td>
</tr>
<tr>
<td>normalized time</td>
<td>−118 ms</td>
</tr>
</tbody>
</table>

$p < 0.001$

$p < 0.01$

**Table**: Estimated differences in the response variables contrasting non-adaptive to adaptive systems. MT metrics calculated by comparing original MT outputs to user corrections.
Summary

- Novel graphical interface with (phrase-) alignments for a new form of interactive post-editing
- Alignment can be used for immediate and bulk adaptation of the translation model
- User study shows significant reductions in manual effort and slight speed improvement

Our code open source: https://github.com/pks/lfpe
Thank you!


References II


