Factored models for phrase-based translation

LING 575 Lecture 5

Kristina Toutanova
MSR & UW

April 25, 2011
With slides mostly borrowed from Philip Koehn
Final project proposals

- Project proposals due May 02
- About two times the work on one of the two-week homework assignments per person
- Project proposal: a short (one or two paragraphs) description of
  - What is the problem and general approach
  - Who is in the group and who will do what
  - What data you are using
Ideas for Final Project Topics-1

- Formulate a new word-alignment model or use new features in an existing model
- Implement a modification to the phrase-table of a phrasal MT system
- Add a new feature to Moses
- Implement a method for mining parallel documents or sentences from comparable corpora or from the web
- Implement extension on syntax-based model using Joshua
  - E.g. refine the nonterminals
Ideas for Final Project Topics-2

- Implement a method for pre or post-processing (e.g. using syntax or morphology)
- Come up with a new method for phrase pair extraction from parallel corpora
- Test a new word-segmentation strategy for MT (e.g. for Chinese or morph. rich languages)
- Experiment with factored models for MT, varying the methods of factor construction and translation decomposition
Alternative Final Project: A Survey

- If you would rather spend more time researching the relevant literature
- Could instead collect and write an overview of three or more related articles
- Also list ideas for future extensions
Overview

- Motivation for factored models
- Example
- Model and Training
- Alternate Decoding Paths
- Decoding
- Applications
  - Enriching output space
  - Translating factored words
  - Enriching input space
Model form

- In standard phrase-based MT we have scores of phrase-pairs
  - \[ \text{score}(f_1 \ldots f_m, e_1 \ldots e_n) = \lambda_1 P(f_1 \ldots f_m | e_1 \ldots e_n) + \lambda_2 P(e_1 \ldots e_n | f_1 \ldots f_m) + \lambda_3 P_{\text{lex}}(f_1 \ldots f_m | e_1 \ldots e_n) + \lambda_4 P_{\text{lex}}(e_1 \ldots e_n | f_1 \ldots f_m) + \lambda_5 \]

- Now the scores of phrase-pairs are decomposed into scores for translation and generation steps within the phrase pair

- Take this model:
Model form equation

\[ f_j = f_j|lf_j|posf_j \]
\[ e_i = e_i|le_i|pose_i|me_i \]

\[
\text{score}(f_1|lf_1|posf_1 \ldots f_m|lf_m|posf_m, e_1|le_1|me_1 \ldots e_n|le_n|pose_n|me_n) = \\
\text{score}(lf_1 \ldots lf_m, le_1 \ldots le_n) + \\
\text{score}(posf_1 \ldots posf_m, pose_1|me_1 \ldots pose_n|me_n) + \\
\text{score}_{gen}(e_1, pose_1|le_1|me_1) + \ldots + \text{score}_{gen}(e_n, pose_n|le_n|me_n)
\]
Specifying factored models in Moses: Example

```
train-factored-phrase-model.perl  --corpus factored-corpus/proj-syndicate.1000  \  --root-dir pos-decomposed  \ 
\  --f de \  --e en  \ 
\  --lm 0:3:factored-corpus/surface.lm:0  \ 
\  --lm 1:3:factored-corpus/pos.lm:0  \ 
\  --translation-factors 0-0  \ 
\  --generation-factors 0-1  \ 
\  --decoding-steps t0,g0
```
Specifying factored models in Moses: Example

```
train-factored-phrase-model.perl ...
--f de --e en \
--lm 0:3:factored-corpus/surface.lm:0 \ 
--lm 2:3:factored-corpus/pos.lm:0 \ 
--translation-factors 1-1+2-2,3\ 
--generation-factors 1,2,3-0 \ 
--decoding-steps t0,t1,g0 \
```
Specifying factored models in Moses: multiple decoding paths

train-factored-phrase-model.perl ...
--f de --e en \
--lm 0:3:factored-corpus/surface.lm:0 \
--lm 2:3:factored-corpus/pos.lm:0 \
--translation-factors 1-1+2-2,3+0-0,2\n--generation-factors 1,2,3-0 \n--decoding-steps t0,t1,g0:t2 \

[Diagram of word, lemma, part-of-speech, and morphology nodes input and output]
Other result on enriching output

[Koehn and Hoang 07]

<table>
<thead>
<tr>
<th>Model</th>
<th>BLEU</th>
</tr>
</thead>
<tbody>
<tr>
<td>baseline (surface)</td>
<td>23.41%</td>
</tr>
<tr>
<td>surface + morph</td>
<td>24.66%</td>
</tr>
<tr>
<td>surface + POS + morph</td>
<td>24.25%</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Model</th>
<th>BLEU</th>
</tr>
</thead>
<tbody>
<tr>
<td>baseline (surface)</td>
<td>25.82%</td>
</tr>
<tr>
<td>surface + all morph</td>
<td>27.04%</td>
</tr>
<tr>
<td>surface + case/number/gender</td>
<td>27.45%</td>
</tr>
<tr>
<td>surface + CNG/verb/prepositions</td>
<td>27.62%</td>
</tr>
</tbody>
</table>

40K training sent

20K training sent
Factored translation models make it possible to model words as a set of features (factors).

We can use this to build pos-based language models for the target:
- Good empirical improvements with 7-gram LMs over output syntactic factors.

We can use this to represent translation of phrases as translation of parts of words in the phrases (e.g. lemma/morphology):
- Using multiple decoding paths we can avoid the strong independence assumptions.
- Good empirical improvement in small/medium data conditions.

We can enrich the word representation of an input language to aid translation into a morphologically richer language:
- Good improvements on specific linguistic phenomena, not a huge boost to overall BLEU.
References

- **Factored Translation Models**, Philipp Koehn and Hieu Hoang, EMNLP 2007, [pdf](#).

- **Enriching Morphologically Poor Languages for Statistical Machine Translation**, Eleftherios Avramidis and Philipp Koehn, ACL 2008, [pdf](#).
Error analysis for an English-Greek baseline phrasal system