Preference Grammars and Soft Syntactic Constraints for GHKM Syntax-based SMT

Matthias Huck, Hieu Hoang, Philipp Koehn

University of Edinburgh

25 October 2014
Introduction

Feature-based integration of syntactic information into GHKM string-to-tree translation

- **Preference grammars:**
  soft target-side syntax
  - Target syntax as a feature rather than via labeled non-terminals in the SCFG

- **Soft syntactic constraints:**
  non-restrictive source-syntactic enhancement
  - No hard source-syntactic constraints (as in standard tree-to-tree translation) imposed in extraction or decoding

Our empirical evaluation: English → German WMT task
Related Work

- **Open-source Moses** implementation for GHKM translation
  - GHKM rule extraction: *Williams and Koehn* (2012)
  - Decoding with CYK+ parsing and cube pruning: *Hoang et al.* (2009)
  - Competitive results for European language pairs: *Nadejde et al.* (2013); *Williams et al.* (2014)
- **Preference grammars**: beneficial as a syntactic extension of hierarchical systems (*Venugopal et al.*, 2009; *Stein et al.*, 2010)
- **Soft syntactic constraints**: related source-syntactic techniques improved hierarchical (*Marton and Resnik*, 2008; *Vilar et al.*, 2008; *Hoang and Koehn*, 2010) and other syntax-based systems (*Zhang et al.*, 2011; *Huang et al.*, 2013) on Chinese→English and Arabic→English tasks
Preference Grammars

- Target-side non-terminals not decorated with syntactic labels, but with a single generic non-terminal symbol

Baseline

\[
\begin{align*}
X, \text{ADJD} & \rightarrow \langle \text{present}, \text{anwesend} \rangle \\
X, \text{ADV} & \rightarrow \langle \text{present}, \text{anwesend} \rangle \\
X, \text{AP-PD} & \rightarrow \langle \text{present}, \text{anwesend} \rangle \\
& \cdots
\end{align*}
\]

Preference grammar system

\[
\begin{align*}
X, X & \rightarrow \langle \text{present}, \text{anwesend} \rangle
\end{align*}
\]

- Distribution of implicit target label vectors stored as additional information with each translation rule

\[
X, X \rightarrow \langle \text{present}, \text{anwesend} \rangle \neq (\text{ADJD}) 0.98 (\text{ADV}) 0.001 (\text{AP-PD}) 0.01 \cdots
\]

- Computation of a tree-wellformedness feature during decoding
Soft Source Syntactic Constraints

- Parse source-side data as well
- GHKM extractor stores an additional rule property: source syntactic label vectors
- Provide parsed input data to the decoder
- During decoding, score matches and mismatches of source label vectors and input labels
- Soft features, no hard constraints

<table>
<thead>
<tr>
<th>rule</th>
<th>source label vectors</th>
</tr>
</thead>
<tbody>
<tr>
<td>$X, VP-OC \rightarrow \langle \text{present, zu präsentieren} \rangle$</td>
<td>$(VB), (NN)$</td>
</tr>
<tr>
<td>$X, ADJD \rightarrow \langle \text{present, anwesend} \rangle$</td>
<td>$(ADJP), (ADVP), \ldots$</td>
</tr>
<tr>
<td>$X, TOP \rightarrow \langle X^0 \text{ is now } X^1 . . \text{jetzt ist } NP-SB^0 \text{ VP-OC}^1 . \rangle$</td>
<td>$(TOP, NP, ADJP), \ldots$</td>
</tr>
<tr>
<td>$X, ^S-TOP \rightarrow \langle \text{is } X^0 X^1, \text{ist } ADV^0 \text{ ADJD}^1 \rangle$</td>
<td>$(VP, ADVP, ADJP), \ldots$</td>
</tr>
</tbody>
</table>
the army is now about to present flowers.
die Armee ist jetzt dabei, Blumen zu präsentieren.
Die Armee ist jetzt dabei, Blumen zu präsentieren.

The army is now about to present flowers.
die Armee ist jetzt dabei, Blumen zu präsentieren.

the army is now about to present flowers.

the army is now present.
Die Armee ist jetzt dabei, Blumen zu präsentieren.

Jetzt ist die Armee zu präsentieren.
jetzt ist die Armee zu präsentieren

the army is now about to present flowers

die Armee ist jetzt dabei, Blumen zu präsentieren

the army is now present
Die Armee ist jetzt dabei, Blumen zu präsentieren.

Jetzt ist die Armee zu präsentieren.
die Armee ist jetzt dabei, Blumen zu präsentieren.

jetzt ist die Armee zu präsentieren.
Präsentieren zu Blumen dabei, ist jetzt die Armee TOP.
jetzt ist die Armee zu präsentieren.
jetzt ist die Armee zu präsentieren.

The army is now about to present flowers.
die Armee ist jetzt dabei, Blumen zu präsentieren.

The army is now about to present flowers.
die Armee ist jetzt dabei, Blumen zu präsentieren.

the army is now about to present flowers.
zu präsentieren

präsentieren
to present

präsentieren

zu präsentieren
to present

präsentieren

to present

anwesend

heute

ADV

ADV

ADV

ADJD
Preference Grammars and Soft Syntactic Constraints for GHKM Syntax-based SMT

Matthias Huck
mhuck@inf.ed.ac.uk

Hieu Hoang
hhoang@inf.ed.ac.uk

Philipp Koehn
pkoehn@inf.ed.ac.uk

Motivation
Feature-based integration of syntactic information into GHKM string-to-tree statistical machine translation
- The hard target-side syntactic constraints that are imposed by the target non-terminal labels might be too restrictive. Should we soften them?

**Preference grammars** promote syntactic well-formedness on the target language side while also allowing for derivations that are not linguistically motivated (as in hierarchical translation)

Soft syntactic constraints augment the system with additional source-side syntax features while not modifying the set of string-to-tree translation rules or the baseline feature scores

Soft syntactic constraints are only defined on the source side, and not on the tree-wellformedness feature.

Preference Grammars

- Target-side non-terminals not decorated with syntactic labels, but with a single generic non-terminal symbol
- Extracted rules which differ only with respect to their non-terminal labels are collapsed to a single entry in the rule table, and their rule counts are pooled
- Distribution of implicit target label vectors stored as additional information with each translation rule
- A source syntactic label vector fully matches the input labels
- Left-hand side non-terminal label mismatch
- Number of right-hand side non-terminals label mismatches

Soft Source Syntactic Constraints

Training:
- Provide syntactic parses of the source side of the data
- GHKM extractor collects the source syntactic labels that cover the source-side span of non-terminals
- Sets of source syntactic label vectors are memorized with the rules as an additional property

Decoding:
- Input data parsed in a preprocessing step
- Computation of three dense features which score matches and mismatches of input labels and source label vectors that are associated with translation rules

Experimental Setup
- English → German WMT task (4.5 M sentence pairs)
- Syntactic annotation: BitPar for German, Berkeley Parser for English
- Right binarization of target parse trees
- SAMT-style composite labels on source side
- Singleton hierarchical rules are discarded
- No more than 50 most frequent label vectors per rule stored
- Decoding with CYK and cube pruning
- Tuning with batch MIRA
- Development set: 2000 selected sentences from newstest2008-2012

**Experimental Results (English → German)**

<table>
<thead>
<tr>
<th>System</th>
<th>dev</th>
<th>newstest2013</th>
<th>newstest2014</th>
</tr>
</thead>
<tbody>
<tr>
<td>GHKM string-to-tree baseline</td>
<td>34.7</td>
<td>47.3</td>
<td>20.0</td>
</tr>
<tr>
<td>Hard source syntactic constraints</td>
<td>34.6</td>
<td>47.4</td>
<td>19.9</td>
</tr>
<tr>
<td>Soft source syntactic constraints</td>
<td>35.1</td>
<td>47.0</td>
<td>20.3</td>
</tr>
<tr>
<td>String-to-string (GHKM syntax-directed extraction)</td>
<td>33.8</td>
<td>48.0</td>
<td>19.3</td>
</tr>
<tr>
<td>Preference grammar</td>
<td>33.9</td>
<td>47.7</td>
<td>19.3</td>
</tr>
<tr>
<td>Soft source syntactic constraints</td>
<td>34.6</td>
<td>47.0</td>
<td>19.8</td>
</tr>
</tbody>
</table>

Sparse Features for Soft Syntactic Constraints

- Large number of binary features which depend on the label identity
- Separate weight tuned for each of them
- Optionally: Restrict the number of sparse features by specifying a core set of labels
- **core = non-composite** – plain constituent labels as given by the syntactic parser (no SAMT-style composite labels)
- **core = dev-min-occ100** – labels in the input data on the development set with minimum occurrence count threshold 100

**Experimental Results (Tuned on newstest2012)**

| System | dev | newstest2012 | newstest2013 | newstest2014 |
|--------|-----|--------------|--------------|
| GHKM string-to-tree baseline | 17.9 | 65.7 | 19.9 | 63.2 |
| Soft source syntactic constraints | 18.2 | 65.3 | 20.3 | 62.6 |
| Sparse features | 18.6 | 64.9 | 20.4 | 62.5 |
| Sparse features (core = non-composite) | 18.4 | 65.1 | 20.3 | 62.7 |
| Sparse features (core = dev-min-occ100) | 18.4 | 64.8 | 20.6 | 62.2 |


References III


