Features for Syntax-based Moses: Project Report

Eleftherios Avramidis, Arefeh Kazemi, Tom Vanallemeersch, Phil Williams, Rico Sennrich, Maria Nadejde, Matthias Huck

MT Marathon

13 September 2014
Features for Syntax-based Moses

Syntax-based translation with Moses

- Phrases as SCFG rules: GHKM minimal and composed rules
- Chart-based decoding
- Core implementation is stable
- Has been applied successfully for multiple language pairs in recent evaluation campaigns

Additional features?

- Some useful features are available for phrase-based Moses, but not for the syntax-based system yet
- Some useful features are implemented in other SMT toolkits, but not in Moses
- Syntax-based translation is attractive for development of novel features: come up with ideas!
Current Project Status

Feature implementation

- Based on Moses’ feature function framework
- Possibly using the new phrase properties framework
- ... Or novel functionality that is not a “feature” in a strict sense

Extending Moses’ GHKM extractor and chart-based decoder with features that are not currently supported

- Sparse feature that combines the NT label with the span length
- Phrase orientation model
- $n$-best tree output, MIRA/MERT with head-word chain metric
- Employ parse trees with semantic information, bilingual tree alignment
- Gender agreement for long distance predicative
- (Morphological smoothing feature)
Span Length (1)

```
S → NP → VBN → NP → JJR → NN → IN → NP → PP → NN
   
NP → VBN → TO → NP → IN
   
VP → MD → VB
   
S → SL = 13
   
VP → SL = 8
   
schools will be encouraged
   
encouraged to focus on greater math, spelling and grammar
```
Span Length (2)

Model span length (SL) of non-terminals:
- For source and target parse tree
- Sparse features
- Computed during decoding

Features:
- LHS label & source SL
- LHS label & target SL
- LHS label & source SL & target SL
- LHS label & (source SL – target SL)
- first 3 features with SL binned (SL/3)

Initial results (cased BLEU):
- Baseline: 22.3
- With span length sparse features: 22.2
Lexicalized reordering for syntax-based translation

Monotone non-terminal orientation

Swap non-terminal orientation

Discontinuous non-terminal orientation

Huck, Wuebker, Rietig, Ney: A Phrase Orientation Model for Hierarchical Machine Translation (WMT 2013)
Phrase Orientation (2)

Moses implementation:

- Four orientation classes, left-to-right + right-to-left direction
- Extract smoothed relative frequencies of orientations given the rule
- Store as additional phrase property

very [X] ||| sehr [ADV] ||| ... ||| {{Orientation 0.72 0.03 0.2 0.05 ...}}

- During decoding, determine orientations and score them with the respective values from the additional property

Prototype available, but not ready for productive use. TODO:

- Handle degenerate cases correctly
- Efficiency
- EMS integration
- Testing
Support for $n$-best tree output

Tree similarity score (HWCM: head-word chain metric) for MERT

Pipeline integration — HWCM requires reference tree instead of reference string

Example:

SRC: he can not get any more .
REF: zu mehr kann er nicht verurteilt werden .

Tree output in $n$-best list:

er kann nicht mehr kriegen . ||| [Q <s> [sent [vroot [subj [PPER er]]] [VMFIN kann] [aux [adv [PTKNEG nicht] [adv [ADV mehr]]] [VVINF kriegen]] [punct $.$]] </s>]

er kann nicht mehr bekommen . ||| [Q <s> [sent [vroot [subj [PPER er]]] [VMFIN kann] [aux [adv [PTKNEG nicht] [adv [ADV mehr]]] [VVINF bekommen]] [punct $.$]] </s>]

er kann nicht mehr aussteigen . ||| [Q <s> [sent [vroot [subj [PPER er]]] [VMFIN kann] [aux [adv [PTKNEG nicht] [adv [ADV mehr]]] [VVINF aussteigen]] [punct $.$]] </s>]

er kann nicht mehr gewöhnen . ||| [Q <s> [sent [vroot [subj [PPER er]]] [VMFIN kann] [aux [adv [PTKNEG nicht] [adv [ADV mehr]]] [VVINF gewöhnen]] [punct $.$]] </s>]

er kann nicht mehr erhalten . ||| [Q <s> [sent [vroot [subj [PPER er]]] [VMFIN kann] [aux [adv [PTKNEG nicht] [adv [ADV mehr]]] [VVINF erhalten]] [punct $.$]] </s>]

Avramidis, Kazemi, Vanallemeersch, Williams, Sennrich, Nadejde, Huck
$n$-best Tree Output,
MIRA/MERT with Head-word Chain Metric (2)
n-best Tree Output, MIRA/MERT with Head-word Chain Metric (3)
Parse Trees with Semantic Information, Bilingual Tree Alignment (1)

We might be mistaken in using the gross domestic product per inhabitant as the sole indicator.

kunnen wij als enige indicator het bruto binnenlands product per hoofd van de bevolking nemen

Avramidis, Kazemi, Vanallemeersch, Williams, Sennrich, Nadejde, Huck

MT Marathon

Features for Syntax-based Moses: Project Report
Parse Trees with Semantic Information, Bilingual Tree Alignment (2)

Avramidis, Kazemi, Vanallemeersch, Williams, Sennrich, Nadejde, Huck

Features for Syntax-based Moses: Project Report
Parse Trees with Semantic Information, Bilingual Tree Alignment (3)

Tree alignment based on:
- GIZA++ lexical probabilities
- semantic role labeling

PropBank:
Arg1 of predicate *use* is the "theme"
Long Distance Agreement (1)

**Problem:** Gender agreement for long distance predicative

**Example:** (for translation to Greek)
The citizens of our countries were victims of natural [disasters:female], which were indeed [terrible:female]

**Feasibility test:** Identify whether there are enough resources

- Get access to the Greek parser
- See whether Greek parser output annotation is sufficient and easily convertible

**Pre-processing of the Greek side**

- Parsed 12,300 sentences... until parser crashed
- Unfortunately parses miss alignment to parallel corpus
- Identified pattern out of dependency tree
TODO:

- Train a syntax-based baseline
- Add the desired constraint

Constraint:

```python
if dependency_label == 'Pnom'
    and pos_tag == 'Aj'
    and pos_tag(dependency.get_parent()) == 'Vb'
    and pos_tag(dependency.get_parent().get_parent()) == 'No':
    node.constrain_gender(dependency.get_parent().get_parent().get_gender())
```
Questions?

Thank you for your attention

Eleftherios Avramidis, Arefeh Kazemi, Tom Vanallemeersch, Phil Williams, Rico Sennrich, Maria Nadejde, Matthias Huck

mhuck@inf.ed.ac.uk