

Step-by-Step

Hieu Hoang
Matthias Huck

December 2014

SMT Pipeline

Preprocessing

- tokenizer
- lowercase

Alignment

Phrase extraction

Tuning

Decoding

Postprocessing

- recasing
- detokenizer

Scoring

- BLEU score

Tokenize and Lowercase

Original

Madam President, on a point of order.

Tokenized

Madam President , on a point of order .

Lowercased

madam president , on a point of order .

Tokenization

- Language-specific
 - Moses tokenizer
 - Basic
 - Supports 22 languages
- Use external tokenizer
 - eg. MADA for Arabic
- Text normalization
- Compound splitter

Casing

- Lowercase / Recase
- Truecase
 - Most common case for each word
- Real case
 - Don't do any case processing
- Dependent on data, language

Word Alignment

Input

frau präsidentin , zur geschäftsordnung .

madam president , on a point of order .

Output

0-0 1-1 2-2 3-3 4-4 4-5 4-6 4-7 5-8

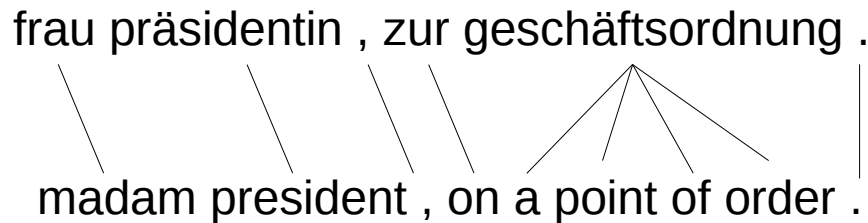
frau präsidentin , zur geschäftsordnung .

madam president , on a point of order .

Phrase Extraction

Input

frau präsidentin , zur geschäftsordnung .
madam president , on a point of order .



Output

frau ||| madam ||| 0-0
präsidentin ||| president ||| 0-0
frau präsidentin ||| madam president ||| 0-0 1-1
.....

Phrase-Table Creation

Input

frau ||| madam ||| 0-0
präsidentin ||| president ||| 0-0
frau prääsidentin ||| madam president ||| 0-0 1-1

.....

Output

frau ||| madam ||| 0.89719 0.852604 0.364141 0.331687 ||| 0-0 ||| 21992 54185 19731 ||| |||
frau ||| her ||| 0.00210656 0.0042994 0.000442927 0.0007843 ||| 0-0 ||| 11393 54185 24
....
präsidentin ||| president ||| 0.163569 0.117671 0.969559 0.865753 ||| 0-0 ||| 89962 15177 14715 ||| |||
präsidentin ||| presided ||| 0.00122554 0.0066225 4.84499e-06 3.74e-05 ||| 0-0 ||| 60 15177 1 ||| |||
....
frau prääsidentin ||| madam president ||| 0.874562 0.100327 0.884875 0.287159 ||| 0-0 1-1 ||| 14844 14671 12982 ||| |||
frau prääsidentin ||| madam chairman ||| 0.933333 0.00801738 0.000954263 0.000359383 ||| 0-0 1-1 ||| 15 14671 14 ||| |||
...

Probabilities

1. $p(\text{source} \mid \text{target})$
2. $p(\text{source} \mid \text{target})$ per word
3. $p(\text{target} \mid \text{source})$
4. $p(\text{target} \mid \text{source})$ per word

Phrase-Table Format

frau ||| madam ||| 0.89719 0.852604 0.364141 0.331687 ||| 0-0 ||| 21992 54185 19731

1.Source

2.Target

3.Scores

4.Word Alignment

5.Counts

- Not used during decoding
- Debugging information

6.Sparse scores

- Key-value pairs
- eg. VB 1 NP 2

7.Key-value properties

- {{Key Values}}
- eg. {{NonTermContext 1 0 23 32 24 51 0.0685714}}

Language Model

Input

Monolingual target text

Output

Standard ARPA format

```
\data\  
ngram 1= 92951  
ngram 2= 3010080  
ngram 3= 14418108  
ngram 4= 29762375  
ngram 5= 40770370
```

```
\1-grams:  
-6.49179 <s> -1.59127  
-4.76751 resumption -0.696029  
....  
\2-grams:  
-5.79014 <s> <s> -0.366199  
-3.99848 <s> resumption -1.76034  
....  
\3-grams:  
-0.279408 <s> <s> <s> 0.114419  
-1.35467 <s> <s> resumption 2.14638  
....
```

moses.ini

Decoder configuration file

input factors

[input-factors]

0

No factors

mapping steps

[mapping]

0 T 0

1 phrase-table (phrase-table 0)

[distortion-limit]

6

Maximum distortion = 6 words

feature functions

[feature]

UnknownWordPenalty

WordPenalty

PhrasePenalty

PhraseDictionaryMemory name=TranslationModel0 num-features=4 path=.../phrase-table.1 input-factor=0 output-factor=0

LexicalReordering name=LexicalReordering0 num-features=6 type=wbe-msd-bidirectional-fe-allff input-factor=0 output-factor=0 path=...

Distortion

KENLM lazyken=1 name=LM0 factor=0 path=.../europarl.binlm.1 order=5

Feature
functions

dense weights for feature functions

[weight]

UnknownWordPenalty0= 1

WordPenalty0= -1

PhrasePenalty0= 0.2

TranslationModel0= 0.2 0.2 0.2 0.2

LexicalReordering0= 0.3 0.3 0.3 0.3 0.3 0.3

Distortion0= 0.3

LM0= 0.5

weights

Running the decoder

- Basic command line
moses -f moses.ini [< input]
- Override moses.ini parameters
moses -f moses.ini -distortion-limit 0
- Short cuts
moses -f moses.ini -dl 0

Hierarchical model

Decoder configuration file

input factors

[input-factors]

0

[search-algorithm]

3

0=standard pb. 1=cube pruning. 3=CYK+

mapping steps

[mapping]

0 T 0

1 T 1

2 phrase-tables. Regular phrase-table + 'glue rules'

[cube-pruning-pop-limit]

1000

Number hypotheses created per stack/cell

[non-terminals]

X

LHS label for rules of unknown words

[max-chart-span]

20

1000

Max span of rules in each phrase table.

20 for regular phrase-table

1000 for glue rules

Hierarchical model

Decoder configuration file

feature functions

[feature]

UnknownWordPenalty

WordPenalty

PhrasePenalty

PhraseDictionaryMemory name=TranslationModel0 num-features=4 path=regular-phrase-table input-factor=0 output-factor=0

PhraseDictionaryMemory name=TranslationModel1 num-features=1 path=glue-grammar input-factor=0 output-factor=0

KENLM lazyken=1 name=LM0 factor=0 path=... order=5

dense weights for feature functions

[weight]

UnknownWordPenalty0= 1

WordPenalty0= -1

PhrasePenalty0= 0.2

TranslationModel0= 0.2 0.2 0.2 0.2

TranslationModel1= 1.0

LM0= 0.5

Tuning

Untuned

[weight]

UnknownWordPenalty0= 1
WordPenalty0= -1
PhrasePenalty0= 0.2
TranslationModel0= 0.2 0.2 0.2 0.2
TranslationModel1= 1.0
LM0= 0.5

Tuned

[weight]

UnknownWordPenalty0= 1
WordPenalty0= -0.336804
PhrasePenalty0= -0.0855363
TranslationModel0= 0.0739741 0.0212178 0.139777 0.0393687
TranslationModel1= 0.17723
LM0= 0.126092

- Iterative process
 - Repeatedly run decoder with different settings
 - Decode held-out tuning data (with reference)
 - 1000-2000 sentences
- Tune on in-domain data
- Multiple algorithms
 - MERT
 - Original
 - Not good for sparse feature
 - PRO
 - MIRA
 - Batch MIRA

Evaluation

- Decode test set
 - 1000-2000 sentences (minimum)
 - With references
 - Multiple references
- Understand metric
 - Test set dependent
 - Relative / absolute
- Many metrics
 - BLEU
 - Nist-BLEU
 - IBM BLEU
 - Multi-BLEU
 - Meteor
 - TER
 -



Step-by-Step

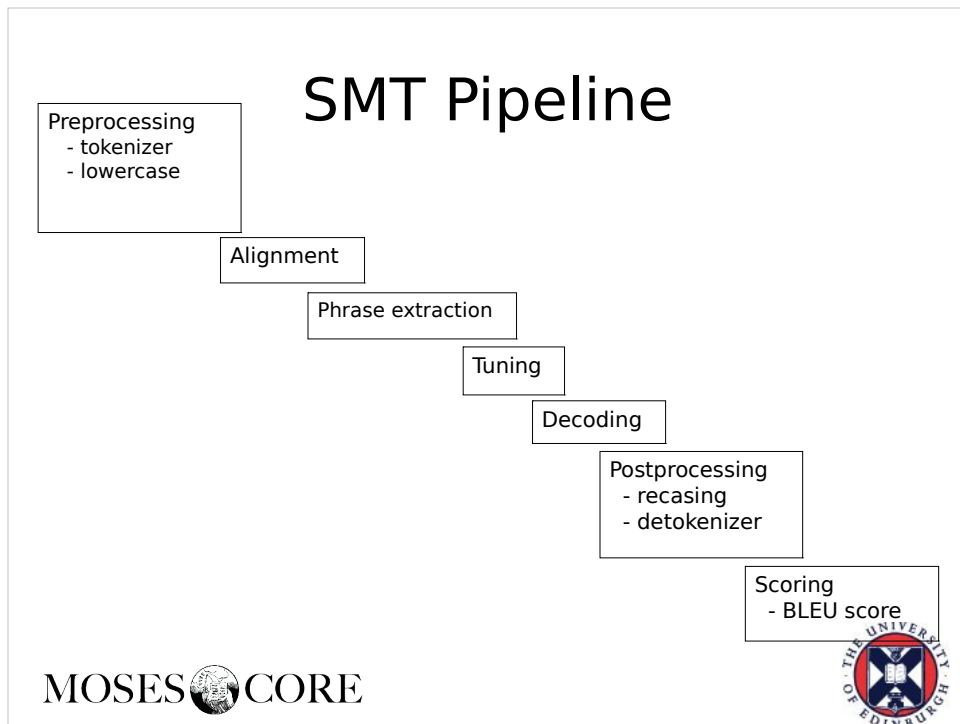
Hieu Hoang
Matthias Huck

December 2014

Thanks for inviting me to come

Here to tell you a little about the things I've
been doing to Moses

- over the past 2 years
- mainly concentrate of the past year
 - but will quickly tell you about things I did
prior to that



What is this feature function framework?

Tokenize and Lowercase

Original

Madam President, on a point of order.

Tokenized

Madam President , on a point of order .

Lowercased

madam president , on a point of order .

Tokenization

- Language-specific
 - Moses tokenizer
 - Basic
 - Supports 22 languages
- Use external tokenizer
 - eg. MADA for Arabic
- Text normalization
- Compound splitter

MOSES  CORE



Casing

- Lowercase / Recase
- Truecase
 - Most common case for each word
- Real case
 - Don't do any case processing
- Dependent on data, language

Word Alignment

Input

frau präsidentin , zur geschäftsordnung .

madam president , on a point of order .

Output

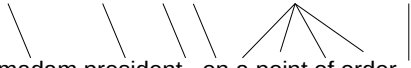
0-0 1-1 2-2 3-3 4-4 4-5 4-6 4-7 5-8

frau präsidentin , zur geschäftsordnung .
| | | | |
madam president , on a point of order .

Phrase Extraction

Input

frau präsidentin , zur geschäftsordnung .
madam president , on a point of order .



Output

frau ||| madam ||| 0-0
präsidentin ||| president ||| 0-0
frau präsidentin ||| madam president ||| 0-0 1-1
.....

Phrase-Table Creation

Input

frau ||| madam ||| 0-0
präsidentin ||| president ||| 0-0
frau präsidentin ||| madam president ||| 0-0 1-1
.....

Output

frau ||| madam ||| 0.89719 0.852604 0.364141 0.331687 ||| 0-0 ||| 21992 54185 19731 ||| |||
frau ||| her ||| 0.00210656 0.0042994 0.000442927 0.0007843 ||| 0-0 ||| 11393 54185 24
.....
präsidentin ||| president ||| 0.163569 0.117671 0.969559 0.865753 ||| 0-0 ||| 89962 15177 14715 ||| |||
präsidentin ||| presided ||| 0.00122554 0.0066225 4.84499e-06 3.74e-05 ||| 0-0 ||| 60 15177 1 ||| |||
.....
frau präsidentin ||| madam president ||| 0.874562 0.100327 0.884875 0.287159 ||| 0-0 1-1 ||| 14844 14671 12982 ||| |||
frau präsidentin ||| madam chairman ||| 0.933333 0.00801738 0.000954263 0.000359383 ||| 0-0 1-1 ||| 15 14671 14 ||| |||
...

Probabilities

1. $p(\text{source} \mid \text{target})$
2. $p(\text{source} \mid \text{target})$ per word
3. $p(\text{target} \mid \text{source})$
4. $p(\text{target} \mid \text{source})$ per word

MOSES  CORE



Phrase-Table Format

frau ||| madam ||| 0.89719 0.852604 0.364141 0.331687 ||| 0-0 ||| 21992 54185 19731

1.Source

2.Target

3.Scores

4.Word Alignment

5.Counts

- Not used during decoding
- Debugging information

6.Sparse scores

- Key-value pairs
- eg. VB 1 NP 2

7.Key-value properties

- {{Key Values}}
- eg. {{NonTermContext 1 0 23 32 24 51 0.0685714}}

MOSES  CORE



Language Model

Input

Monolingual target text

Output

Standard ARPA format

```
\data\
ngram 1= 92951
ngram 2= 3010080
ngram 3= 14418108
ngram 4= 29762375
ngram 5= 40770370

....

\1-grams:
-6.49179 <s> -1.59127
-4.76751 resumption -0.696029
....

\2-grams:
-5.79014 <s> <s> -0.366199
-3.99848 <s> resumption -1.76034
....

\3-grams:
-0.279408 <s> <s> <s> 0.114419
-1.35467 <s> <s> resumption 2.14638
....
```

MOSES  CORE



moses.ini

Decoder configuration file

```
# input factors
[input-factors]
0

# mapping steps
[mapping]
0 T 0

[distortion-limit]
6

# feature functions
[feature]
UnknownWordPenalty
WordPenalty
PhrasePenalty
PhraseDictionaryMemory name=TranslationModel0 num-features=4 path=.../phrase-table.1 input-factor=0 output-factor=0
LexicalReordering name=LexicalReordering0 num-features=6 type=wbe-msd-bidirectional-fe-allf input-factor=0 output-factor=0 path=...
Distortion
KENLM lazyken=1 name=LM0 factor=0 path=.../europarl.binlm.1 order=5

# dense weights for feature functions
[weight]
UnknownWordPenalty0= 1
WordPenalty0= -1
PhrasePenalty0= 0.2
TranslationModel0= 0.2 0.2 0.2 0.2
LexicalReordering0= 0.3 0.3 0.3 0.3 0.3 0.3
Distortion0= 0.3
LM0= 0.5
```

No factors

1 phrase-table (phrase-table 0)

Maximum distortion = 6 words

Feature functions

weights

MOSES  CORE



Running the decoder

- Basic command line
moses -f moses.ini [< input]
- Override moses.ini parameters
moses -f moses.ini -distortion-limit 0
- Short cuts
moses -f moses.ini -dl 0

Hierarchical model

Decoder configuration file

```
# input factors
[input-factors]
0

[search-algorithm]
3
# mapping steps
[mapping]
0 T 0
1 T 1

[cube-pruning-pop-limit]
1000

[non-terminals]
x

[max-chart-span]
20
1000
```

0=standard pb. 1=cube pruning. 3=CYK+

2 phrase-tables. Regular phrase-table + 'glue rules'

Number hypotheses created per stack/cell

LHS label for rules of unknown words

Max span of rules in each phrase table.
20 for regular phrase-table
1000 for glue rules

Hierarchical model

Decoder configuration file

```
# feature functions
[feature]
UnknownWordPenalty
WordPenalty
PhrasePenalty
PhraseDictionaryMemory name=TranslationModel0 num-features=4 path=regular-phrase-table input-factor=0 output-factor=0
PhraseDictionaryMemory name=TranslationModel1 num-features=1 path=glue-grammar input-factor=0 output-factor=0
KENLM lazyken=1 name=LM0 factor=0 path=... order=5

# dense weights for feature functions
[weight]
UnknownWordPenalty0= 1
WordPenalty0= -1
PhrasePenalty0= 0.2
TranslationModel0= 0.2 0.2 0.2 0.2
TranslationModel1= 1.0
LM0= 0.5
```

MOSES  CORE



Tuning

Untuned

[weight]

UnknownWordPenalty0= 1
WordPenalty0= -1
PhrasePenalty0= 0.2
TranslationModel0= 0.2 0.2 0.2 0.2
TranslationModel1= 1.0
LM0= 0.5

Tuned

[weight]

UnknownWordPenalty0= 1
WordPenalty0= -0.336804
PhrasePenalty0= -0.0855363
TranslationModel0= 0.0739741 0.0212178 0.139777 0.0393687
TranslationModel1= 0.17723
LM0= 0.126092

- Iterative process
 - Repeatedly run decoder with different settings
 - Decode held-out tuning data (with reference)
 - 1000-2000 sentences
- Tune on in-domain data
- Multiple algorithms
 - MERT
 - Original
 - Not good for sparse feature
 - PRO
 - MIRA
 - Batch MIRA

MOSES  CORE



Evaluation

- Decode test set
 - 1000-2000 sentences (minimum)
 - With references
 - Multiple references
- Understand metric
 - Test set dependent
 - Relative / absolute
- Many metrics
 - BLEU
 - Nist-BLEU
 - IBM BLEU
 - Multi-BLEU
 - Meteor
 - TER
 -

MOSES  CORE

