Feature Functions

Hieu Hoang
Matthias Huck

December 2014

Feature Function

- Calculate score(s) for a translation rule
  - Partial translation
  - Completed translation
- Examples
  - Phrase table
  - Language model
  - Word penalty
  - Phrase penalty
- Many feature functions
  - Weighted linear combination
- What is a translation?
  - Made of multiple ***translation rules***

MOSES CORE

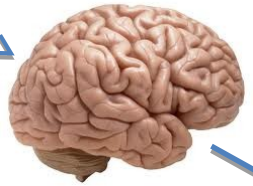
THE UNIVERSITY OF EDINBURGH

Timeline of a Translation Rule

File

Load

Memory



Apply to input sentence

Translation Option

Search

Hypothesis

MOSES CORE

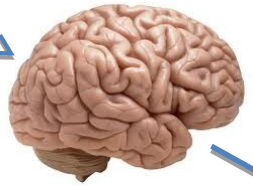
Timeline of a Translation Rule

File

Load

Source phrase
Target phrase

Memory



Apply to input sentence

Input sentence
Input path

Translation Option

Search

Translation context
Segmentation

Hypothesis

MOSES CORE

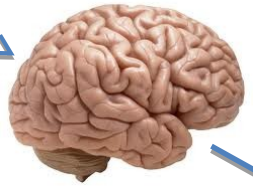
Timeline of a Translation Rule

File

Load

Once

Memory



Apply to input sentence

Per occurrence in sentence

Translation Option

Search

Per hypothesis

Hypothesis

MOSES CORE

Feature Function API

Loading

File

Memory

je suis ||| I am

Access to: Source phrase: je suis

Target phrase: I am

```
void EvaluateInIsolation(
            source,
            target,
            scores,
            estimated future scores)
```

Feature functions that use this:

- Word Penalty
- Phrase penalty
- Language model (partial)

MOSES CORE

THE UNIVERSITY OF EDINBURGH

Feature Function API

Apply to input sentence

Memory

Access to: Input sentence: je suis 25 ans .

Input subphrase: je suis 25

```
void EvaluateWithSourceContext(
        input,
        input path,
        target,
        scores,
        estimated future scores)
```

Translation Option

Feature functions that use this:

- Input feature
- Bag-of-word features....

MOSES CORE

THE UNIVERSITY OF EDINBURGH

Feature Function API

Search

Translation Option

Hypothesis

Access to: Current rule (hypothesis)
Previous rules
Segmentation

Stateful features:

```
state EvaluateWhenApplied(hypo,
        previous state,
        scores)
state EvaluateWhenApplied(hypo
          previous state,
          scores)
```

Stateless features:

```
void EvaluateWhenApplied(hypo, scores)
void EvaluateWhenApplied(hypo, scores)
```

MOSES CORE

THE UNIVERSITY OF EDINBURGH

Feature Function API

Decoding

Translation Option

Hypothesis

Feature functions that use this:

- All stateful features
  - Language models
  - Distortion model
  - Lexicalized reordering model
  - …

MOSES CORE

THE UNIVERSITY OF EDINBURGH

Feature Function

Loading:

```
void Evaluate(source,
              target,
                  scores,
              estimated future scores)
```

Apply to Input:

```
void Evaluate(input,
        input path,
                scores)
```

Search:

Stateful features:

```
state Evaluate(hypo,
        previous state,
        scores)
state EvaluateChart(hypo
          previous state,
          scores)
```

Stateless features:

```
void Evaluate(hypo, scores)
void EvaluateChart(hypo, scores)
```

MOSES CORE

Strange Features functions (1)

- Language model
  - implement 2 Evaluate()

1. Loading
   - evaluate full n-grams

     reprise de la session ||| resumption of the session
   - estimate future cost
     - partial n-grams
2. Search
   - evaluate overlapping n-grams

MOSES CORE

THE UNIVERSITY OF EDINBURGH

Strange Feature Functions (2)

- Phrase-tables
- Unknown Word Penalty
- Generation Model
  - integral part of decoding process
  - Uses no Evaluate()
    - scores assign by decoder

MOSES CORE

THE UNIVERSITY OF EDINBURGH

Extensions

- Change input
  - Add/delete word
  - Integrate parser/tagger
- Prune
  - Hard constraint
  - Negative infinity score
  - Need positive weight
    - Make FF non-tuneable

MOSES CORE

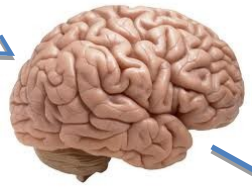
THE UNIVERSITY OF EDINBURGH

Future Extensions

File

Load

Memory



Apply to input sentence

Translation Option

Evaluate, conditioned on other translation option

Search

Hypothesis

MOSES CORE

Feature Functions

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

Feature Function

- Calculate score(s) for a translation rule
  - Partial translation
  - Completed translation
- Examples
  - Phrase table
  - Language model
  - Word penalty
  - Phrase penalty
- Many feature functions
  - Weighted linear combination
- What is a translation?
  - Made of multiple ***translation rules***

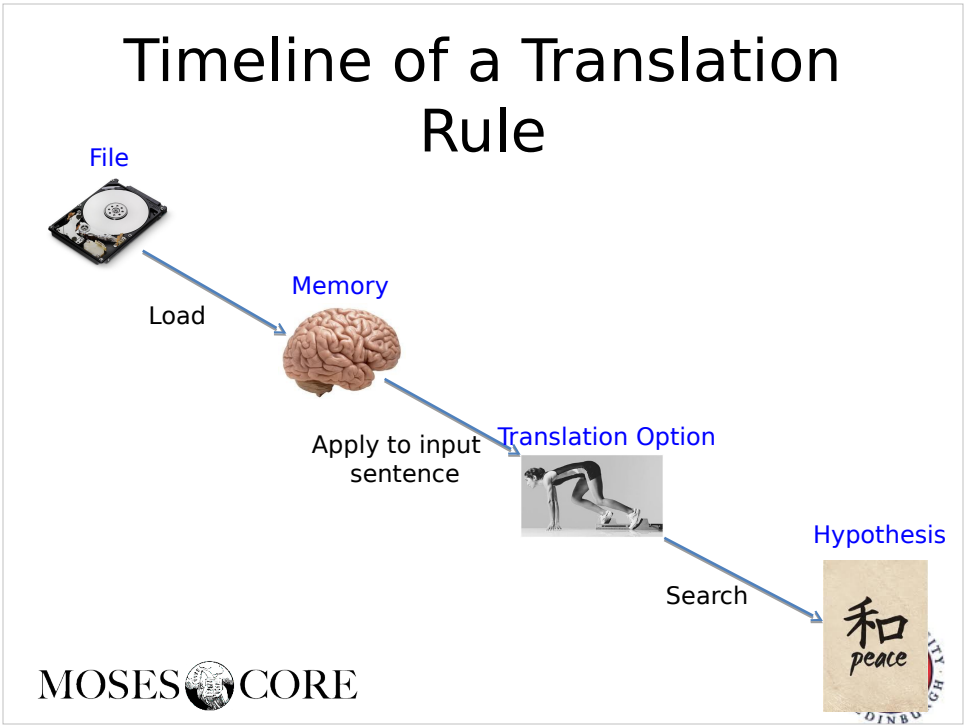
MOSES CORE

What is the task of a feature function
- it's task is to give score to a translation rule
  - the thing you see in a phrase-table
- FF calculate 1 or more scores

These are examples of FF

There can be many scores

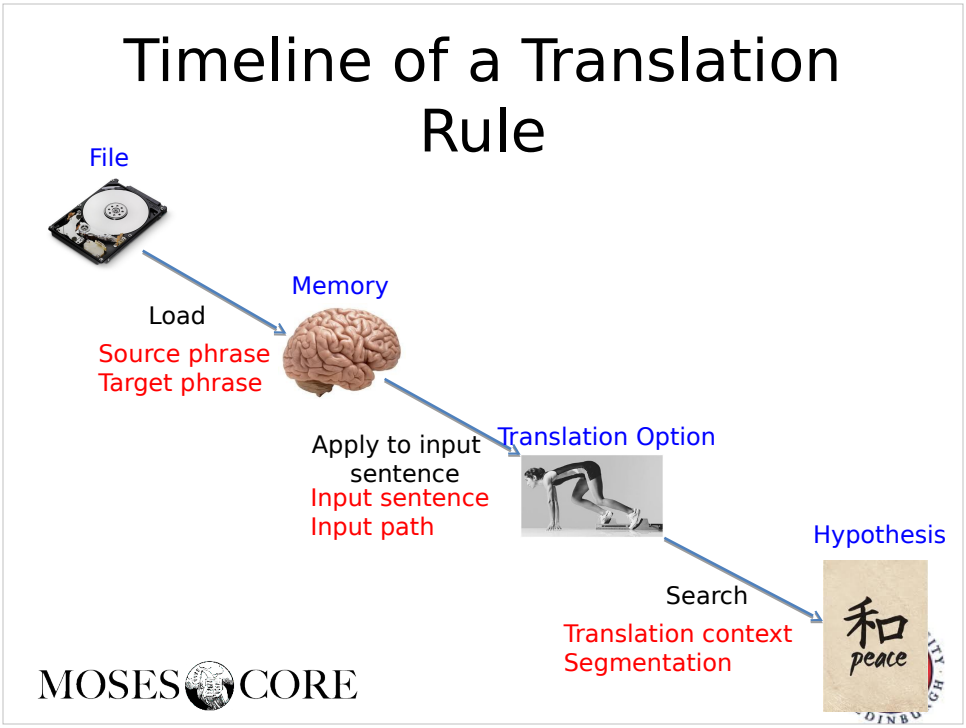
Total score for the translation
- weighted sum of all scores



Translation rule has a lifespan
- starts off in a file on disk
- gets loaded into memory

- before a sentence is decoded
  - translation rules are looked up
    - fitted to a specific place in the source sentence
  - name of translation rules
    - changes to translation option
    - all intents and purposes
      - is a trans rule

When it's being used in search
- it's name changes to hypothesis
  - again, wrapper around a translation rule



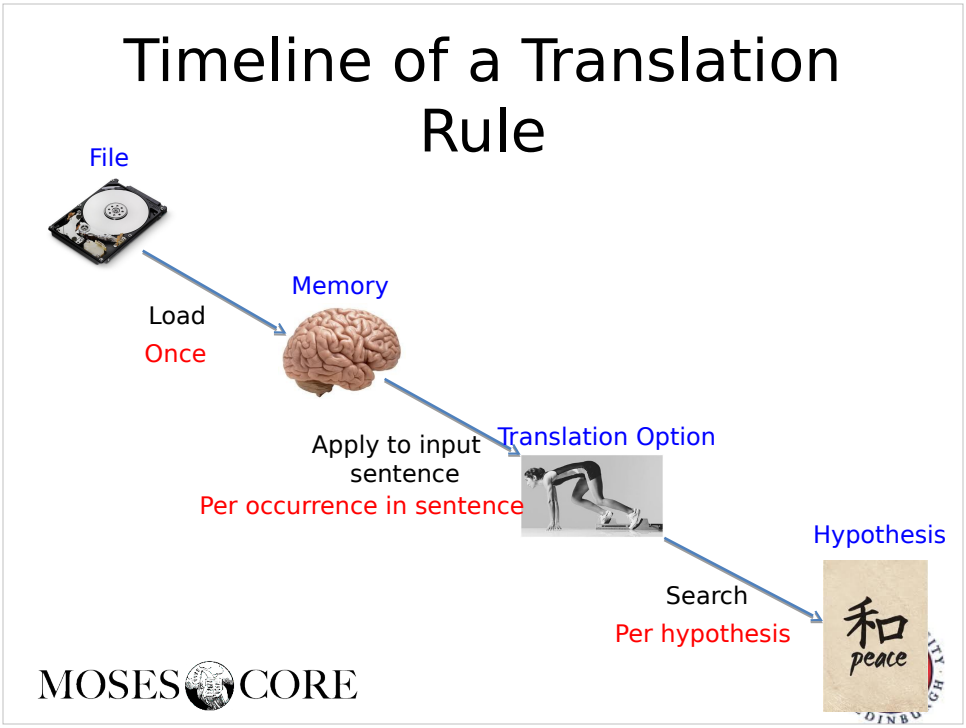
at each step
- feature function has access to different kinds of side information
  with which to score the rule

During loading
- only know what the rule is, without context

When it is being applied to a sentence
- it know the sentence

During search
- it know what other rules have been used

These are the information it can use to score



Point of showing you this timeline

sooner you caclulate it, the better

1. efficiency
   - not repeated
2. more accurate
   - each stage – subject to pruning
     - some rules are thrown away
     - if the feature function can give a good score
     - the rule can say
       ‘hey I’ll be really useful to you, don’t throw me away!”

Feature Function API

Loading

File

Memory

je suis ||| I am

Access to: Source phrase: je suis

Target phrase: I am

```
void EvaluateInIsolation(
        source,
        target,
        scores,
        estimated future scores)
```

Feature functions that use this:

- Word Penalty
- Phrase penalty
- Language model (partial)

MOSES CORE

During load

- this is the translation rule

- If you want your FF to score the rule now
    - implement this function
    - it takes are arguments
        - source + target parts of the rule
        - you return the scores and estimated future score

Feature Function API

Apply to input sentence

Memory

Access to: Input sentence: je suis 25 ans .
Input subphrase: je suis 25

```
void EvaluateWithSourceContext(
        input,
        input path,
        target,
        scores,
        estimated future scores)
```

Feature functions that use this:
- Input feature
- Bag-of-word features....

Translation Option

MOSES CORE

Have a sentence
- looking up rules that can be used in that sentence
- once you find a rule that can be applied
  - to a specific substring in a specific sentence
  - create translation option

At this point
- have another opportunity to evaluate the scores of the rules

Feature Function API

Search

Translation Option

Hypothesis

Access to: Current rule (hypothesis)
Previous rules
Segmentation

Stateful features:

```
state EvaluateWhenApplied(hypo,
        previous state,
        scores)
state EvaluateWhenApplied(hypo
          previous state,
          scores)
```

Stateless features:

```
void EvaluateWhenApplied(hypo, scores)
void EvaluateWhenApplied(hypo, scores)
```

MOSES CORE

search

when you have a translation rule
- you know exactly where it's going to be applied to
- and you actually apply it

Implement 1 of these functions

Only place where calculating the feature function is different for phrase-based or syntax models
- slightly different for stateless and stateful features

This function has access to the hypothesis

Feature Function API

Decoding

Translation Option

Hypothesis

Feature functions that use this:

- All stateful features
  - Language models
  - Distortion model
  - Lexicalized reordering model
  - ...

MOSES CORE

all the translation rules that were used, the total output phrase segmentation

- derivation tree if hiero/syntax model

Feature Function

Loading:

```
void Evaluate(source,
              target,
                  scores,
              estimated future scores)
```

Apply to Input:

```
void Evaluate(input,
        input path,
                scores)
```

Search:

Stateful features:

```
state Evaluate(hypo,
        previous state,
        scores)
state EvaluateChart(hypo
          previous state,
          scores)
```

Stateless features:

```
void Evaluate(hypo, scores)
void EvaluateChart(hypo, scores)
```

MOSES CORE

Recap

- you can score translation rule at 3 stages in the decoding process
  - Loading
  - Appying to the input sentence
  - Search
- Implement 1 of these functions if you do

However, a FF can score the same rule in more than 1 stage

- ie. It can implement more of these functions

Strange Features functions (1)

- Language model
  - implement 2 Evaluate()
  1. Loading
     - evaluate full n-grams

       reprise de la session ||| resumption of the session
     - estimate future cost
       - partial n-grams
  2. Search
     - evaluate overlapping n-grams

MOSES CORE

For those who know Moses

- this is nothing new
- this is the way Moses has always computed language model scores
  - if you had a trigram LM
  - store trigrams in the target phrase upon loading
  - store overlapping n-gram during search

- the new framework enable this optimisation to be used by every other feature function

Strange Feature Functions (2)

- Phrase-tables
- Unknown Word Penalty
- Generation Model
  - integral part of decoding process
  - Uses no Evaluate()
    - scores assign by decoder

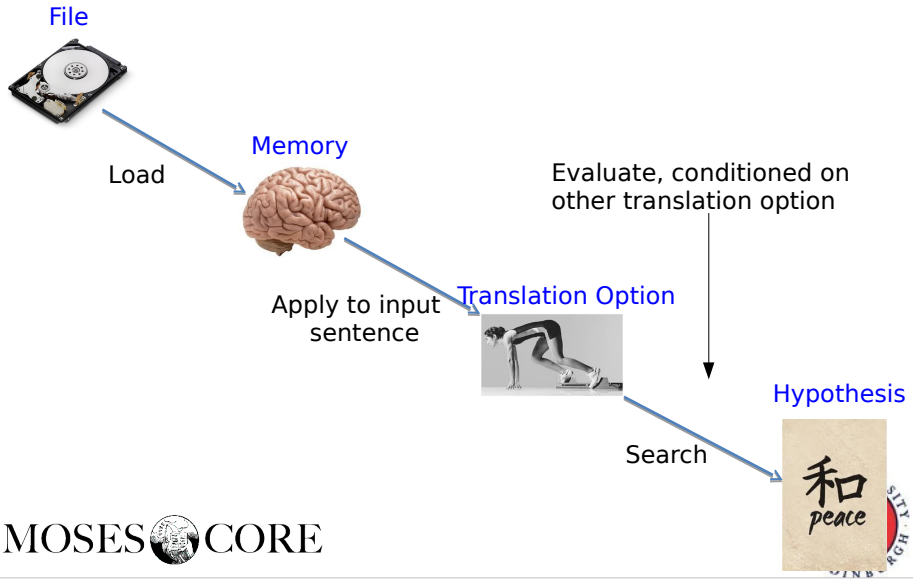
MOSES CORE

Extensions

- Change input
  - Add/delete word
  - Integrate parser/tagger
- Prune
  - Hard constraint
  - Negative infinity score
  - Need positive weight
    - Make FF non-tuneable

MOSES CORE

Future Extensions



MOSES CORE