Decoding by Dynamic Chunking for Statistical Machine Translation

Authors:
Sirvan Yahyaei
Christof Monz

Presenter:
Austin Matthews
Types of Translation Models

- Phrase-Based
  - No long-distance reordering
- Syntax-Based
  - Very restrictive
Types of Translation Models

- No long-distance reordering
- Phrase-Based
- Chunking!
- Syntax-Based

Very restrictive
Previous Chunking Attempts

- POS-Based (Zhang et. al 2007)
  - Requires tagger per language

- Source reordering (Xia and McCord 2004)
  - Makes irreversible decisions early
  - Lessens effect of LM!

- Instead, let the decoder make these decisions
What’s in a Chunk?

- Similar to phrase extraction
What’s in a Chunk?

- Similar to phrase extraction

|   | $f_1$ | $f_2$ | $f_3$ | $f_4$ | $f_5$ | $f_6$ | $f_7$ |
|---|-------|-------|-------|-------|-------|-------|-------|
| $e_1$ |      |       |       |       |       |       |       |
| $e_2$ |      |       |       |       |       |       |       |
| $e_3$ |      |       |       |       |       |       |       |
| $e_4$ |      |       |       |       |       |       |       |
| $e_5$ |      |       |       |       |       |       |       |
| $e_6$ |      |       |       |       |       |       |       |
| $e_7$ |      |       |       |       |       |       |       |
What’s in a Chunk?

- Similar to phrase extraction
What’s in a Chunk?

- Similar to phrase extraction
Scoring Chunks

- MaxEnt Classifier, tuned with L-BFGS
- Features: Number of times a chunk boundary occurred ...
  - before $f_j$
  - after $f_j$
  - between $f_j$ and $f_{j+1}$
  - weight of $(f_j, f_{j+1})$
Integration with the Decoder

Feature: \[ h_{chunk}(f^J_1, e^I_1, C, S) = \log \prod_{1}^{J} (C_j S(j) + (1 - C_j)(1 - S(j))) \]

- Allow reordering chunks ONLY
  - No local reordering!
- Each chunk is translated monotonically Left → Right or Right → Left
- Compete iff same covered words & same last chunked position
Decoding Operations

- Insert chunk boundary
- Translate Phrase
- Reorder chunks
A Few More Parameters

- Max Chunk Length
- Min # of Chunks
- Max # of Chunks
- Distortion Limit!
A Few More Parameters

- Max Chunk Length
- Min # of Chunks
- Max # of Chunks
- Distortion Limit!

- In addition to stack size, beam width, phrase length limit, target phrases per source phrase
Start state:

[ man muss die schwierigkeiten bei der bestimmung von ursache und wirkung anerkennen . ]

Insert chunk boundary:

[ man muss ][ die schwierigkeiten bei der bestimmung von ursache und wirkung anerkennen . ]
Decoding: Example

Insert chunk boundary:

[ man muss ][ die schwierigkeiten bei der bestimmung von ursache und wirkung anerkennen . ]

Translate phrase:

[ man muss ][ die schwierigkeiten bei der bestimmung von ursache und wirkung anerkennen . ]

we must
Translate phrase:

[man muss][ die schwierigkeiten bei der bestimmung von ursache und wirkung anerkennen . ]

we must

Insert chunk boundary:

[man muss][ die schwierigkeiten bei der bestimmung von ursache und wirkung ][ anerkennen . ]

we must
Decoding: Example

Insert chunk boundary:

[man muss ][ die schwierigkeiten bei der bestimmung von ursache und wirkung ][ anerkennen . ]
we must

Insert chunk boundary:

[man muss ][ die schwierigkeiten bei der bestimmung von ursache und wirkung ][ anerkennen ][ . ]
we must
Insert chunk boundary:

[man muss][ die schwierigkeiten bei der bestimmung von ursache und wirkung][ anerkennen ][ . ]

we must

Reorder:

[man muss][ anerkennen ][ die schwierigkeiten bei der bestimmung von ursache und wirkung ][ . ]

we must
Decoding: Example

Reorder:

[man muss ][ anerkennen ][ die schwierigkeiten bei der bestimmung von ursache und wirkung ][ . ]

we must

Translate phrase:

[man muss ][ anerkennen ][ die schwierigkeiten bei der bestimmung von ursache und wirkung ][ . ]

we must recognize
Translate phrase:

we must recognize

Translate phrase:

we must recognize the difficulties in the provision of cause and effect
Translate phrase:

[man muss][ anerkennen ][ die schwierigkeiten bei der bestimmung von ursache und wirkung ][ . ]

we must recognize the difficulties in the provision of cause and effect.

Translate phrase:

[man muss][ anerkennen ][ die schwierigkeiten bei der bestimmung von ursache und wirkung ][ . ]

we must recognize the difficulties in the provision of cause and effect.
Experimental Setup

- DE → EN system trained on Europarl and News Commentary (~1.4M sentences / 40M words)
- Tested on EP and NC test sets separately (~2k sentences each)
- Standard phrase-based baseline
### Results

|   | Test Set | System   | BLEU  | NIST   | TER  |
|---|----------|----------|-------|--------|------|
| 1 | EP       | Baseline | 26.76 | 7.0063 | 66.26|
| 2 | EP       | Chunk    | 27.16 | 7.1084 | 67.39|
| 3 | NC       | Baseline | 24.54 | 7.1591 | 65.24|
| 4 | NC       | Chunk    | 24.87 | 7.1798 | 64.01|

- MaxEnt classifier error rate independantly tested with 73% accuracy.
- During decoding, 32% of chunk boundaries differed from MaxEnt’s predictions.
My Thoughts

- Interesting intermediate between syntax-based and pure phrasal
- Monotonically translating chunks seems too restrictive
- Perhaps not much to gain if extracted phrases are good
Discussion