Decoding by Dynamic Chunking for Statistical Machine Translation

Sirvan Yahyaei & Christof Monz
MT Summit 2009

Presented by Nguyen Bach, Feb 24, 2010

In a nutshell

- Contributions:
  - A phrase-based decoder that dynamically chunks and translates the source sentences
- Claims
  - Chunking classifier has no language specific and syntactic information
  - Enable the decoder to make long distance reordering

What is a chunk-based decoder?

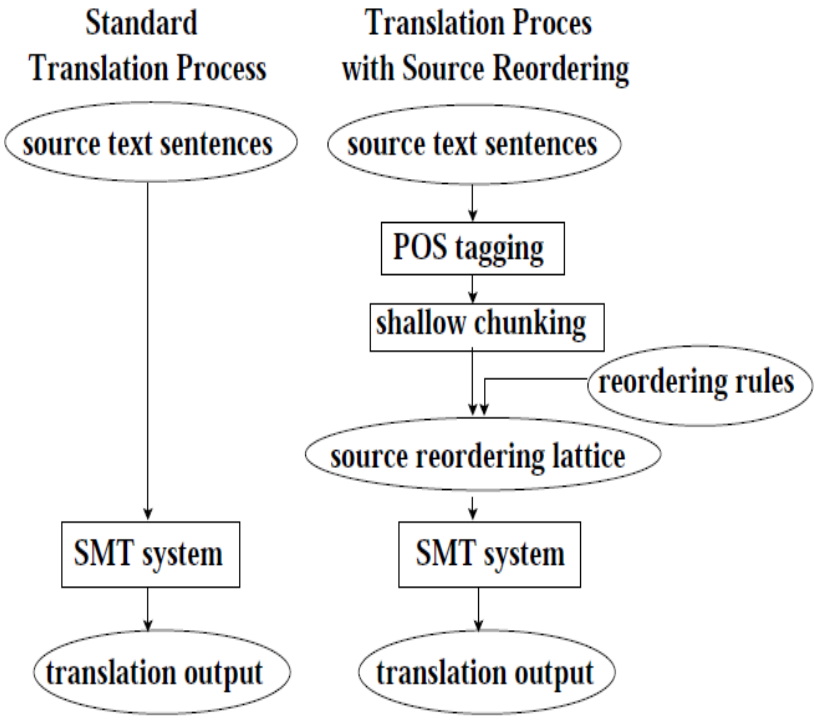

Figure 1: Illustration of the translation process with and without source reordering.

Zhang, Zen and Ney (SSST 2007)

- Syntactic chunk: NP &VP
- Rules: chunk + POS tags

Table 1: Examples of reordering rules. (lhs: chunk and POS tag sequence, rhs: permutation)

no.	lhs	rhs
1.	$NP_0\ PP_1\ u_2\ n_3$	0 1 2 3
2.	$NP_0\ PP_1\ u_2\ n_3$	3 0 1 2
3.	$DNP_0\ NP_1\ VP_2$	0 1 2
4.	$DNP_0\ NP_1\ VP_2$	1 0 2
5.	$DNP_0\ NP_1\ m_2$	0 1 2
6.	$DNP_0\ NP_1\ m_2\ ad_3$	3 0 1 2
7.	$DNP_0\ NP_1\ m_2\ ad_3\ v_4$	4 3 0 1 2

What are they doing?

- Chunking & Decoding at the same time
  - Decisions at each level (chunking, chunk-based re-ordering, and translation) are not made independently of each other.
- Chunk Scorer: a binary classifier
  - Classify every point between 2 source words into two classes: ‘chunk boundary’ and ‘no chunk boundary’
- Decoding by Chunking
  - Step 1: chunking
  - Step 2: continue chunking or translating monotonically a chunk by phrase-based system

Chunk Scorer

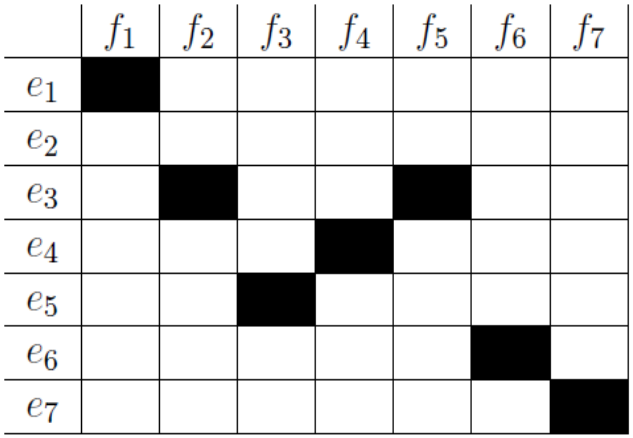

Figure 1: An example of chunks with left to right, (f_1, f_2), (f_6, f_7) and right to left (f_3, f_4, f_5) orientations.

- A chunk: a contiguous group of words that can be translated monotonically from left to right or right to left.
- A chunk boundary between f_j and j_{j+1} if there is no source word aligned to $\{i \mid a_j < i < a_{j+1}\}$
  - No chunk boundary: f_6 & f_7 ; f_1 &f_2
  - Exist a boundary: f_2 & f_3

- MaxEnt features: h1-being a chunk boundary or not, h2- left word of a chunk, h3-right word, h4-significance of pair in data

Decoding by Chunking

1	[man muss] die schwierigkeiten bei der bestimmung von ursache und wirkung anerkennen .
C	
2	[**man muss**] die schwierigkeiten bei der bestimmung von ursache und wirkung anerkennen .
P	we must
3	[man muss][die schwierigkeiten bei der bestimmung von ursache und wirkung] anerkennen .
C	we must
4	[man muss][die schwierigkeiten bei der bestimmung von ursache und wirkung][anerkennen] .
C	we must
5	[man muss][die schwierigkeiten bei der bestimmung von ursache und wirkung][**anerkennen**] .
P	we must recognise
6	[man muss][**die schwierigkeiten bei der bestimmung von ursache und wirkung**][anerkennen] .
P	we must recognise the difficulties in the provision of cause and effect

Decoding by Chunking

- Hypothesis recombination:
  - 2 hyps can be combined if they have identical chunk boundaries for uncovered positions
  - Source word coverage vector, LM history
- Decoder features
  - Chunking cost: cost from maxent classifier
  - Chunking penalty: reward/penalizes each chunking applications based on the sign of its weight
  - Chunk distortion model: similar to distance-based DM

Experiments

- Chunk classifier: 73% accuracy
  - LM & TM scores also contribute to chunking decision
  - 32% cases classifier says Yes but decoder says No

	Run	System	BLEU	NIST	1−TER
1	EP	Baseline	0.2687	7.0063	0.3374
2	EP	Chunk	0.2716	7.1084	0.3261
3	NC	Baseline	0.2454	7.1591	0.3476
4	NC	Chunk	0.2487	7.1798	0.3599

Table 2: Results on German to English task of ACL WMT 2008 translation task, Europarl (EP) and News Commentary (NC) test sets. Since TER is measuring the error, 1−TER is reported. Default values are used for parameters of the chunking decoder (see 3.3).

3 main reasons for the chunking system to fail

- A wrong classification decision by the chunking scorer may lead the decoder to jump or monotonically translate in a wrong position
- The classifier picks a proper chunking boundary, the other features force the decoder to apply the wrong re-ordering.
- Even with accurate chunk boundaries, the decoder can still fail to apply the correct re-orderings

Are their claims verified?

- Chunking classifier has no language specific and syntactic information
  - YES
- Enable the decoder to make long distance reordering
  - Not very clear: no analysis on average jump distance