Approaches to Machine Translation: Rule-based, Statistical and Hybrid

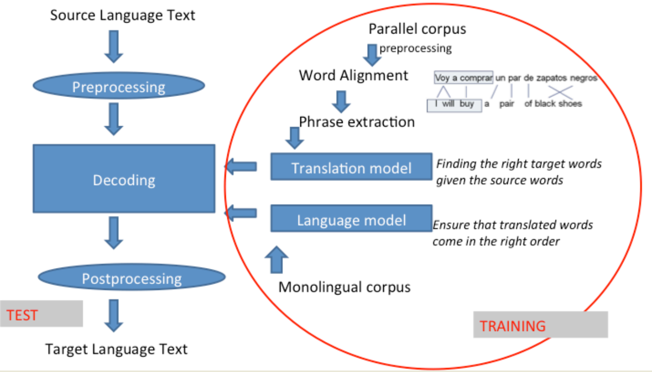
Growing heuristics, Symmetrization and Phrase-Models

Lluís Formiga

July 22, 2014

UNIVERSITAT POLITÈCNICA DE CATALUNYA
BARCELONATECH

A picture is worth a million equations



Summary

- Lexical Translation / Alignment and Reordering
- Phrase Model
- Training the model
    - word alignment
    - phrase pair extraction
    - phrase pair scoring
- Log linear model
    - sub-models as feature functions
    - lexical weighting
    - word and phrase count features
- EM training of the phrase model

Talk Outline

Symmetrization

Alignment Heuristics

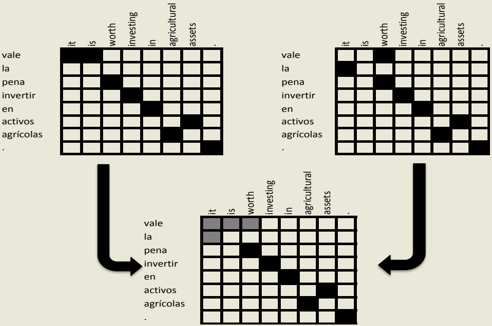
Moving from words to phrases

Motivations

Model

Reordering Models

Symmetrizing Word Alignments



- Intersection of GIZA++ bidirectional alignments
- Grow additional alignment points [Och and Ney, CompLing2003]

Growing heuristic

grow-diag-final(e2f,f2e)

neighboring = {(-1,0),(0,-1),(1,0),(0,1),(-1,-1),(-1,1),(1,-1),(1,1)}
alignment A = intersect(e2f,f2e); grow-diag(); final(e2f); final(f2e);

grow-diag()

while new points added **do**
  for all Target Language word $e \in [1 \dots e_n]$, Source Language word $f \in [1 \dots f_n]$, $(e, f) \in A$ **do**
    for all neighboring alignment points (e_{new}, f_{new}) **do**
      if (e_{new} unaligned OR f_{new} unaligned) AND $(e_{new}, f_{new}) \in$ union(e2f,f2e) **then**
        add (e_{new}, f_{new}) to A
      end if
    end for
  end for
end while

Growing heuristic

```
final()
for all Target Language word e_new ∈ [1...e_n], Source Language word f_new ∈ [1...f_n] do
    if (e_new unaligned OR f_new unaligned) AND (e_new, f_new) ∈ union(e2f,f2e) then
        add (e_new, f_new) to A
    end if
end for
```

Growing heuristic with FINAL-AND

grow-diag-final-AND(e2f,f2e)

```
neighboring = {(-1,0),(0,-1),(1,0),(0,1),(-1,-1),(-1,1),(1,-1),(1,1)}
alignment A = intersect(e2f,f2e); grow-diag(); final-AND(e2f); final-AND(f2e);
```

final-AND()

```
for all Target Language word e_new ∈ [1...e_n], Source Language word f_new ∈ [1...f_n] do
    if (e_new unaligned AND f_new unaligned) AND (e_new, f_new) ∈ union(e2f,f2e) then
        add (e_new, f_new) to A
    end if
end for
```

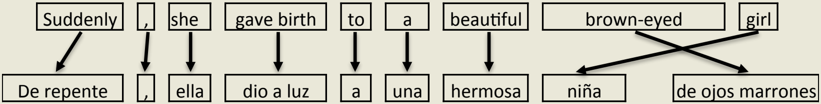
Discriminative Training Methods

- Given some annotated training data, supervised learning methods are possible
- Structured prediction
  - not just a classification problem
  - solution structure has to be constructed in steps
- Many approaches: maximum entropy, neural networks, support vector machines, conditional random fields, MIRA, ...
- Small labeled corpus may be used for parameter tuning of unsupervised aligner [Fraser and Marcu, 2007]

Moving From Words to Phrases

- Word-Based Models translate words as atomic units
- Phrase-Based Models translate phrases as atomic units
- Advantages:
  - many-to-many translation can handle non-compositional phrases
  - use of local context in translation
  - the more data, the longer phrases can be learned
- ``Standard Model'', used by Google Translate and others

Phrase-Based Model



- Foreign input is segmented in phrases
- Each phrase is translated into Target
- Phrases are reordered

Phrase Translation Table

- Main knowledge source: table with phrase translations and their probabilities
- Example: phrase translations for suddenly

Translation	**Probability** $\phi(\bar{\mathsf{e}}\|\bar{\mathsf{f}})$
de repente	0.27
repentinamente	0.12
de pronto	0.09
súbitamente	0.05

Real Example

- Phrase translations for den Vorschlag learned from the Europarl corpus:

English	$\phi(\bar{e}\|\bar{f})$	English	$\phi(\bar{e}\|\bar{f})$
the proposal	0.6227	the suggestions	0.0114
's proposal	0.1068	the proposed	0.0114
a proposal	0.0341	the motion	0.0091
the idea	0.0250	the idea of	0.0091
this proposal	0.0227	the proposal ,	0.0068
proposal	0.0205	its proposal	0.0068
of the proposal	0.0159	it	0.0068
the proposals	0.0159	...	...

  - lexical variation (proposal vs suggestions)
  - morphological variation (proposal vs proposals)
  - included function words (the, a, ...)
  - noise (it)

Linguistic Phrases?

- Model is not limited to linguistic phrases (noun phrases, verb phrases, prepositional phrases, ...)
- Example non-linguistic phrase pair

brown - eyed $\longrightarrow$ de ojos marrones

- Prior noun often helps with translation of preposition
- Experiments show that limitation to linguistic phrases hurts quality

Probabilistic Model

- Bayes rule

$$\begin{aligned} e_{\text{best}} &= \text{argmax}_e \, p(e|f) \\ &= \text{argmax}_e \, p(f|e) \, p_{\mathrm{LM}}(e) \end{aligned}$$

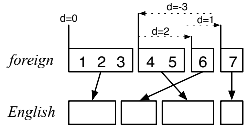
  - translation model $p(e|f)$
  - language model $p_{\mathrm{LM}}(e)$
- Decomposition of the translation model

$$p(\bar{f}_1^I|\bar{e}_1^I) = \prod_{i=1}^{I} \phi(\bar{f}_i|\bar{e}_i) \, d(\text{start}_i - \text{end}_{i-1} - 1)$$

  - phrase translation probability ϕ
  - reordering probability d

Distance-Based Reordering

- Distance-Based
- Lexical-Based (week 3)



phrase	translates	movement	distance
1	1–3	start at beginning	0
2	6	skip over 4–5	+2
3	4–5	move back over 4–6	-3
4	7	skip over 6	+1

Scoring function:

$d(x) = \alpha^{|x|}$ — exponential with distance

Talk Outline

Symmetrization

- Alignment Heuristics

Moving from words to phrases

- Motivations
- Model
- Reordering Models