Data Intensive Linguistics — Lecture 16
Machine translation (III): Decoding

Philipp Koehn
6 March 2006

**Statistical Machine Translation**
- Components: Translation model, language model, decoder

**Phrase-Based Translation**
- Foreign input is segmented in phrases
  - any sequence of words, not necessarily linguistically motivated
- Each phrase is translated into English
- Phrases are reordered

**Phrase Translation Table**

| English          | φ(e|f) | English          | φ(e|f) |
|------------------|-------|------------------|-------|
| the proposal      | 0.6227| the suggestions  | 0.0114|
| a proposal        | 0.1088| the proposed     | 0.0114|
| the idea          | 0.0341| the motion       | 0.0091|
| this proposal     | 0.0207| its proposal     | 0.0068|
| of the proposal   | 0.0159| its proposal     | 0.0068|
| the proposals     | 0.0159| ...              | ...   |

**Decoding Process**
- Build translation left to right
  - select foreign words to be translated
- One to many translation
Decoding Process

- Many to one translation

Translation Options

- Look up possible phrase translations
  - many different ways to segment words into phrases
  - many different ways to translate each phrase

Hypothesis Expansion

- Start with empty hypothesis
  - e: no English words
  - f: no foreign words covered
  - p: probability 1

A Quick Word on Probabilities

- Not going into detail here, but...

- Translation Model
  - phrase translation probability $p(Mary|\text{Maria})$
  - reordering costs
  - phrase/word count costs
  - ...

- Language Model
  - uses trigrams:
  - $p(\text{Mary did not}) = p(\text{Mary|START}) \times p(\text{did|Mary,START}) \times p(\text{not|Mary did})$
Hypothesis Expansion

- Add another hypothesis
- Further hypothesis expansion

Hypothesis Expansion

- ... until all foreign words covered
  - find best hypothesis that covers all foreign words
  - backtrack to read off translation

Explosion of Search Space

- Number of hypotheses is exponential with respect to sentence length
  ⇒ Decoding is NP-complete [Knight, 1999]
  ⇒ Need to reduce search space
    - risk free: hypothesis recombination
    - risky: histogram/threshold pruning

Hypothesis Recombination

- Different paths to the same partial translation
  ⇒ Combine paths
    - drop weaker path
    - keep pointer from weaker path (for lattice generation)

- Recombined hypotheses do not have to match completely
  - No matter what is added, weaker path can be dropped, if:
    - last two English words match (matters for language model)
    - foreign word coverage vectors match (effects future path)
Hypothesis Recombination

- Recombined hypotheses do not have to match completely
- No matter what is added, weaker path can be dropped, if:
  - last two English words match (matters for language model)
  - foreign word coverage vectors match (effects future path)
⇒ Combine paths

Hypothesis Stacks

- Organization of hypotheses into stacks
  - here: based on number of foreign words translated
  - during translation all hypotheses from one stack are expanded
  - expanded hypotheses are placed into stacks

Future Cost Estimation

- Estimate cost to translate remaining part of input
- Step 1: estimate future cost for each translation option
  - look up translation model cost
  - estimate language model cost (no prior context)
  - ignore reordering model cost
  → LM * TM = p(to) * p(the|to) * p(to the|a la)

Future Cost Estimation: Step 3

- Step 3: find cheapest future cost path for each span
  - can be done efficiently by dynamic programming
  - future cost for every span can be pre-computed

Future Cost Estimation: Application

- Use future cost estimates when pruning hypotheses
- For each uncovered contiguous span:
  - look up future costs for each maximal contiguous uncovered span
  - add to actually accumulated cost for translation option for pruning

Future Cost Estimation: Step 2

- Step 2: find cheapest cost among translation options

Comparing Hypotheses

- Comparing hypotheses with same number of foreign words covered
  - Λ: Maria no dio una bofetada a la bruja verde
  - better partial translation
  - covers easier part → lower cost

Pruning

- Hypothesis recombination is not sufficient
  ⇒ Heuristically discard weak hypotheses early
- Organize Hypothesis in stacks, e.g. by
  - same foreign words covered
  - same number of foreign words covered
  - same number of English words produced
- Compare hypotheses in stacks, discard bad ones
  - histogram pruning: keep top n hypotheses in each stack (e.g., n=100)
  - threshold pruning: keep hypotheses that are at most α times the cost of best hypothesis in stack (e.g., α = 0.001)
A* search

- Pruning might drop hypothesis that lead to the best path (search error)

- A* search: safe pruning
  - future cost estimates have to be accurate or underestimates
  - lower bound for probability is established early by depth first search: compute cost for one complete translation
  - if cost-so-far and future cost are worse than lower bound, hypothesis can be safely discarded

- Not commonly done, since not aggressive enough

Limits on Reordering

- Reordering may be limited
  - Monotone Translation: No reordering at all
  - Only phrase movements of at most n words

- Reordering limits speed up search (polynomial instead of exponential)
- Current reordering models are weak, so limits improve translation quality

Word Lattice Generation

- Search graph can be easily converted into a word lattice
  - can be further mined for n-best lists
  - enables reranking approaches
  - enables discriminative training

Sample N-Best List

- Simple N-best list:

| Translation || Reordering | LM | TM | Word Penalty || Score |
|-------------|-------------|----|----|-------------|--------|
| this is a small house || -7 -28.5107 -2.52573 -5 || -38.0364 |
| it is a small house || -7 -28.7683 -2.52573 -5 || -38.2940 |
| this is an small house || -7 -28.8999 -1.83258 -5 || -37.7325 |
| it is an small house || -7 -28.3018 -3.91202 -5 || -39.2139 |
| this is a house small || -7 -28.0443 -3.91202 -5 || -38.9563 |
| it is a house small || -7 -28.3018 -3.21888 -5 || -39.2139 |
| this is a house little || -7 -28.5689 -3.21888 -5 || -39.4015 |
| it is a house little || -7 -28.8557 -3.91202 -5 || -39.0566 |
| this house is a little || -7 -28.0443 -3.91202 -5 || -38.9563 |
| it 's a little house || -7 -28.5689 -3.91202 -5 || -39.4015 |
| this house is a small || -7 -28.3018 -3.91202 -5 || -39.2139 |
| it is a small house || -7 -28.7683 -3.91202 -5 || -38.2940 |
| this is a little house || -7 -28.5107 -3.91202 -5 || -38.0364 |
| it is a little house || -7 -28.7683 -2.52573 -5 || -38.2940 |
| this house is a small || -7 -28.0443 -2.52573 -5 || -38.9563 |
| it is a small house || -7 -28.3018 -2.52573 -5 || -39.2139 |
| this is a small house || -7 -28.5107 -2.52573 -5 || -38.0364 |
| it is a small house || -7 -28.7683 -2.52573 -5 || -38.2940 |

XML Markup

Er erzielte <NUMBER english='17.55'>17,55</NUMBER> Punkte .

- Add additional translation options
  - number translation
  - name translation

- Additional options
  - provide multiple translations
  - provide probability distribution along with translations
  - allow bypassing of provided translations