

11-731: Machine Translation

Homework Assignment #5:

Out: Monday, Feb 14th, 2011 (happy valentine :-))
Due: Beginning of class Monday, Feb 21st, 2011
(email thuylinh@cs.cmu.edu)

In this homework you will manually generate the translation lattice and run the decoding for a given sentence.

Part 1: Manual Monotone Phrase-Based Decoding

- i) Use the phrase table provided below to draw the translation lattice for the sentence

el comunicado de la casa blanca

Phrase Table:

blanca # white # 0.7
casa # house # 0.6
casa # home # 0.3
casa blanca # white house # 0.3
comunicado # message # 0.7
comunicado # communication # 0.1
comunicado de # message from # 0.1
de # of # 0.3
de # from # 0.2
de la # from the # 0.11
de la # of the # 0.2
el # the # 0.1
la # the # 0.9
la casa # the house # 0.1

Each line contains a source phrase, a target phrase and just one probability.

- ii) Find the best translation using monotone decoding (i.e. no reordering). As discussed in the class, perform the calculations in negative log domain (i.e. as costs). For each node in the lattice, list the partial hypotheses present up to that node. Also give the score for each hypothesis.
- iii) Add a phrase count feature to the phrase table by appending a constant cost of e^{-1} (i.e. 0.3679 or in negative log domain), to each phrase pair.

Ex: blanca # white # 0.7 0.3679

Now repeat (ii), but this time use the sum of the two feature values.

- iv) Do you see different results for (ii) and (iii)? Explain what you think has happened.

Part 2: Lexicalized Distortion Model

An entry in a lexical reordering table has the following format:

```
source_phrase ||| target_phrase ||| fm fs fd bm bs bd
```

where:

fm: forward monotone prob

fs: forward swap prob

fd: forward discontinuous prob

bm: backward monotone prob

bs: backward swap prob

bd: backward discontinuous prob

You are given the following lexical reordering table:

```
f1 ||| e1 ||| 0.6 0.2 0.2 0.6 0.2 0.2
```

```
f2 ||| e2 ||| 0.1 0.7 0.2 0.6 0.2 0.2
```

```
f3 ||| e3 ||| 0.4 0.2 0.4 0.3 0.3 0.3
```

Show step-by-step how to calculate the lexical reordering probability for the following translations:

i) f1 f2 f3 --> e1 e2 e3

ii) f1 f2 f3 --> e2 e1 e3