

Machine Translation
WiSe 2015/2016



IBM Models – Exercise 3

Dr. Mariana Neves

November 9th, 2015

Goal

- Implement word-based IBM Models

Exercises

- Model 1 (mandatory)
 - Deadline together with exercise 2 (Nov, 22nd)

Algorithm for Model 1

Input: set of sentence pairs (\mathbf{e}, \mathbf{f})

Output: translation prob. $t(e|f)$

```

1: initialize  $t(e|f)$  uniformly
2: while not converged do
3:   // initialize
4:    $\text{count}(e|f) = 0$  for all  $e, f$ 
5:    $\text{total}(f) = 0$  for all  $f$ 
6:   for all sentence pairs  $(\mathbf{e}, \mathbf{f})$  do
7:     // compute normalization
8:     for all words  $e$  in  $\mathbf{e}$  do
9:        $\text{s-total}(e) = 0$ 
10:      for all words  $f$  in  $\mathbf{f}$  do
11:         $\text{s-total}(e) += t(e|f)$ 
12:      end for
13:    end for
14:    // collect counts
15:    for all words  $e$  in  $\mathbf{e}$  do
16:      for all words  $f$  in  $\mathbf{f}$  do
17:         $\text{count}(e|f) += \frac{t(e|f)}{\text{s-total}(e)}$ 
18:         $\text{total}(f) += \frac{t(e|f)}{\text{s-total}(e)}$ 
19:      end for
20:    end for
21:  end for
22:  // estimate probabilities
23:  for all foreign words  $f$  do
24:    for all English words  $e$  do
25:       $t(e|f) = \frac{\text{count}(e|f)}{\text{total}(f)}$ 
26:    end for
27:  end for
28: end while

```

Exercises

- Model 2 (optional)
 - 0.5 point in the exam's grade (20 points)
 - Deadline together with exercise 2 (Nov, 22nd)
- Model 3 (optional)
 - 1.0 point in the exam's grade (20 points)
 - No deadline, until the last week of the course (Jan, 31st), but should also include presentation.
- Algorithms for Models 2 and 3 are presented in the book

Submission

- Deadline (Model 1, mandatory)
 - Sunday, Nov 22rd, 23:59
- Hand-in
 - SQL file with queries
 - Stored procedures in the team's schema
- Presentation
 - Monday, Nov 23rd (together with exercise 2)
- Not graded, but all teams need to present at some point