METL – TP3
IBM Models 1 & 2

Alexandre Kabbach
alexandre.kabbach@unige.ch

Paola Merlo
paola.merlo@unige.ch

19.03.2020

Evaluation: You are allowed an unlimited number of submissions in order to receive feedback. When you are satisfied with your work, you can ask for it to be graded. You can also ask for it to be graded upon a single submission, without receiving feedback. All your TPs must have been graded and must have received an average grade of at least 4/6 for you to register for the METL exam. Indicative deadline: April 1 2020 (this TP should take you two weeks).

1 IBM Model 1

das

e	$t(e\|f)$
the	0.7
that	0.15
which	0.075
who	0.05
this	0.025

Haus

e	$t(e\|f)$
house	0.8
building	0.16
home	0.02
household	0.015
shell	0.005

ist

e	$t(e\|f)$
is	0.8
's	0.16
exists	0.02
has	0.015
are	0.005

klein

e	$t(e\|f)$
small	0.4
little	0.4
short	0.1
minor	0.06
petty	0.04

Given the translation tables above, compute explicitly the translation probability for each of the following translations of the German sentence *das Haus ist klein* (assume the most likely alignment under IBM Model 1):

1. the house is small
2. the house is little
3. small house the is
4. the

Answer the following questions:

a. Is the IBM Model 1 by itself a good model for finding the best translation?

b. Explain how the noisy-channel model compensates for some of the problems of IBM Model 1 as a translation model.

2 IBM Alignment Models 1 & 2

Consider the following two translations of the German sentence g: *natürlich ist das Haus klein* and the associated alignments:

1. e_1 = *of course the house is small*; $a_1 = [1, 1, 3, 4, 2, 5]$
2. e_2 = *the course of small is house*; $a_2 = [3, 1, 1, 5, 2, 4]$

According to IBM Model 1, what are $p(e_1, a_1|g)$ and $p(e_2, a_2|g)$? Are these translation probabilities necessarily equal? And according to IBM Model 2? Explain.

3 Most Probable Alignment in IBM Model 2

The solution to $a^\star = \arg\max_a p(a|f, e)$ can be efficiently found in the IBM Alignment Model 1 since alignment probabilities are all equal and lexical translations are independent of one another. Can the most probable alignment according to IBM Alignment Model 2 also be computed efficiently? Discuss and write pseudo-code illustrating the most efficient solution you can devise. *Hint:* show first that $\arg\max_a p(a|f, e) = \arg\max_a p(f, a|e)$.