CS4LL5 Advanced Computational Linguistics

Aplying brute force EM on IBM model 1

Séamus Woods
15317173

11/11/2019

0.1 Iteration 1

From iteration 1, we have:

$tr(o\|s)$	gr	ho	the
ca	$\frac{1}{2}$	$\frac{1}{2}$	$\frac{1}{2}$
ve	$\frac{1}{2}$	$\frac{1}{4}$	0
la	0	$\frac{1}{4}$	$\frac{1}{2}$

Our task is to complete two further iterations.

0.2 Iteration 2

Calculating $num(a_n^n)$...

$num(a_1^1) = \frac{1}{2} * \frac{1}{4} = \frac{1}{8}$

$num(a_2^1) = \frac{1}{2} * \frac{1}{2} = \frac{1}{4}$

$num(a_1^2) = \frac{1}{2} * \frac{1}{2} = \frac{1}{4}$

$num(a_2^2) = \frac{1}{2} * \frac{1}{4} = \frac{1}{8}$

From this we can calculate our conditional probabilities $P(a_n^d|o, s)$, we need to sum the $num(a^n)$ by summing across the table and use it as the denominator.

$$P(a_1^1|o_1, s_1) = \frac{\frac{1}{8}}{(\frac{1}{8}+\frac{1}{4})} = \frac{1}{3}$$

$$P(a_2^1|o_1, s_1) = \frac{\frac{1}{4}}{(\frac{1}{8}+\frac{1}{4})} = \frac{2}{3}$$

$$P(a_1^2|o_2, s_2) = \frac{\frac{1}{4}}{(\frac{1}{4}+\frac{1}{8})} = \frac{2}{3}$$

$P(a_2^2|o_2, s_2) = \frac{\frac{1}{8}}{(\frac{1}{4}+\frac{1}{8})} = \frac{1}{3}$

Now we make a count by going through all the alignments and incrementing the count by how many times o is paired with s in the alignment and multiplying that by the above conditional alignment probabilities.
We can use the table given to us by the spec, which tells us the counts already.

cnt	gr	ho	the
ca	$1 * \frac{1}{3}$	$2 * \frac{2}{3}$	$1 * \frac{1}{3}$
ve	$1 * \frac{2}{3}$	$1 * \frac{1}{3}$	0
la	0	$1 * \frac{1}{3}$	$1 * \frac{2}{3}$

and for these counts we get $tr(o|s)$ by normalising by column sums:

$P(o\|s)$	gr	ho	the
ca	$\frac{1}{3}$	$\frac{2}{3}$	$\frac{1}{3}$
ve	$\frac{2}{3}$	$\frac{1}{6}$	0
la	0	$\frac{1}{6}$	$\frac{2}{3}$

This is the end of the second iteration, now we'll start the third.

0.3 Iteration 3

Again, we calculate $num(a_n^n)$...

$num(a_1^1) = \frac{1}{3} * \frac{1}{6} = \frac{1}{18}$

$num(a_2^1) = \frac{2}{3} * \frac{2}{3} = \frac{4}{9}$

$num(a_1^2) = \frac{2}{3} * \frac{2}{3} = \frac{4}{9}$

$num(a_2^2) = \frac{1}{3} * \frac{1}{6} = \frac{1}{18}$

Again, we calculate conditional probabilities..

$P(a_1^1|o_1, s_1) = \frac{\frac{1}{18}}{(\frac{1}{18}+\frac{4}{9})} = \frac{1}{9}$

$P(a_2^1|o_1, s_1) = \frac{\frac{4}{9}}{(\frac{1}{18}+\frac{4}{9})} = \frac{8}{9}$

$P(a_1^2|o_2, s_2) = \frac{\frac{4}{9}}{(\frac{1}{18}+\frac{4}{9})} = \frac{8}{9}$

$P(a_2^2|o_2, s_2) = \frac{\frac{1}{18}}{(\frac{1}{18}+\frac{4}{9})} = \frac{1}{9}$

Count multiplied by conditional probabilities...

cnt	gr	ho	the
ca	$1 * \frac{1}{9}$	$2 * \frac{8}{9}$	$1 * \frac{1}{9}$
ve	$1 * \frac{8}{9}$	$1 * \frac{1}{9}$	0
la	0	$1 * \frac{1}{9}$	$1 * \frac{8}{9}$

and for these counts we get $tr(o|s)$ by normalising by column sums:

$P(o\|s)$	gr	ho	the
ca	$\frac{1}{9}$	$\frac{8}{9}$	$\frac{1}{9}$
ve	$\frac{8}{9}$	$\frac{1}{18}$	0
la	0	$\frac{1}{18}$	$\frac{8}{9}$

This is the end of the third iteration.

0.4 Conclusion

History over the 3 iterations.

Obs	Src	0	1	2	3
green	casa	$\frac{1}{3}$	$\frac{1}{2}$	$\frac{1}{3}$	$\frac{1}{9}$
house	casa	$\frac{1}{3}$	$\frac{1}{2}$	$\frac{2}{3}$	$\frac{8}{9}$
the	casa	$\frac{1}{3}$	$\frac{1}{2}$	$\frac{1}{3}$	$\frac{1}{9}$
green	verde	$\frac{1}{3}$	$\frac{1}{2}$	$\frac{2}{3}$	$\frac{8}{9}$
house	verde	$\frac{1}{3}$	$\frac{1}{4}$	$\frac{1}{6}$	$\frac{1}{18}$
the	verde	$\frac{1}{3}$	0	0	0
green	la	$\frac{1}{3}$	0	0	0
house	la	$\frac{1}{3}$	$\frac{1}{4}$	$\frac{1}{6}$	$\frac{1}{18}$
the	la	$\frac{1}{3}$	$\frac{1}{2}$	$\frac{2}{3}$	$\frac{8}{9}$

Hopefully it's clear from the table that the $tr(o|s)$ should converge to:

$P(o\|s)$	gr	ho	the
ca	0	1	0
ve	1	0	0
la	0	0	1