Assignment: applying brute force EM on IBM model 1

October 18, 2019

There were some notes showing a worked example of brute force EM training with IBM model 1. This assignment varies the example a bit.

Suppose a corpus of 3 pairs:

| $s^1$ | green house |
| $o^1$ | casa verde |
| $s^2$ | the house  |
| $o^2$ | la casa    |

Below the 1st iteration of the brute force EM algorithm is shown. You need to complete the 2nd and 3rd iteration.

To apply the brute force EM algorithm, for each pair, each of its possible alignments has to be considered. In the real IBM Model 1, a possibility is to align positions in $o$ with NULL, and in the worked example we excluded that. In this assignment we going to restrict the possibilities even further: the alignments will be 1-to-1, and not possibly n-to-1: so if $i$ is the value $a(j)$ for some $j$ it cannot be the value for any other $j'$. This means there are just 2 possibilities for each pair:

- $a_1^1$, $a_2^1$ for the 2 possible alignments between $o^1$ and $s^1$
- $a_1^2$, $a_2^2$ for the 2 possible alignments between $o^2$ and $s^2$

The table below shows translations probabilities, with $tr(o|s)$ shown at row $o$, col $s$, and they are all initialised to $\frac{1}{3}$

\[
\begin{array}{c|ccc}
   & gr & ho & the \\
---&-----&-----&-----
ca & $\frac{1}{3}$ & $\frac{1}{3}$ & $\frac{1}{3}$ \\
ve & $\frac{1}{3}$ & $\frac{1}{3}$ & $\frac{1}{3}$ \\
la & $\frac{1}{3}$ & $\frac{1}{3}$ & $\frac{1}{3}$ \\
\end{array}
\]

To execute the brute force EM algorithm we need first for the pairs $o^1, s^1$ and $o^2, s^2$ to determine the conditional alignment probabilities so $p(a|o^1, s^1)$ and $p(a|o^2, s^2)$. The slides gave a derivation of the formula for $p(a|o, s)$, it came out to be

\[
p(a|o, s) = \frac{\prod_j [p(o_j|s_{a(j)})]}{\sum_{a'} \prod_j [p(o_j|s_{a'(j)})]} \quad (1)
\]

and in the derivation $\frac{1}{(s^1 + 1)(s^2 + 1)}$ terms cancelled. Even though for this assignment we are restricting the space of assignments the same formula applies – its applicability just depends on the value $p(a|l_o, s)$ being a uniform distribution\(^1\).

As a name for the numerator term in (1) we will use $num(a)$

So to determine the $p(a|o, s)$ values for each pair we need to

---

\(^1\)Numerically for each $a$ the term is $\frac{1}{s^1}$
1. for each possible \( a \) determine \( \text{num}(a) \) (ie. \( \prod_j [p(o_j|s_{a(j)})] \))

2. sum these to give the denominator \( \sum_a \text{num}(a) \) and then take ratios

Armed with these conditional probabilities can then compute expected counts of \( o, s \) combinations across the corpus, and from these recalculate \( \text{tr}(o|s) \) probabilities.

**ITERATION 1**

considering the first pair, for each \( a^1_n \) calculate \( \text{num}(a^1_n) \):

\[
\text{num}(a^1_1) = \frac{1}{9}, \quad \text{num}(a^1_2) = \frac{1}{9}
\]

ditto

sim. for each \( a^2_n \) calculate \( \text{num}(a^2_n) \). At this stage, these all work out as \( \frac{1}{9} \).

from these to calculate the conditional probabilities \( P(a^d_n|o, s) \), need to sum the \( \text{num}(a^n) \) by summing across the table and use it as denominator ie.

\[
P(a^d_n|o^d, s^d) = \frac{\text{num}(a^d_n)}{\sum_n \text{num}(a^d_n)}
\]

\[
P(a^1_1|o^1, s^1) = \frac{1}{9}, \quad P(a^1_2|o^1, s^1)
\]

\[
= \frac{1}{9}, \quad \text{ditto}
\]

\[
P(a^2_1|o^2, s^2) = \frac{1}{9}, \quad P(a^2_2|o^2, s^2)
\]

\[
= \frac{1}{9}, \quad \text{ditto}
\]

Notice these numbers make intuitive sense: with all \( \text{tr}(o|s) \) set equal, all alignments should be equally probable, giving a value of \( \frac{1}{9} \) for each.

Now for each possible vocabulary combination \( o, s \) combination we have to 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.

For these short sentences the \( o, s \) count for any alignment is at most 1, and it will be handy for the calculations to note for each \( (o, s) \) the alignments where it occurs once\(^2\):

<table>
<thead>
<tr>
<th></th>
<th>gr</th>
<th>ho</th>
<th>the</th>
</tr>
</thead>
<tbody>
<tr>
<td>ca</td>
<td>1:1</td>
<td>1:2</td>
<td>1:-</td>
</tr>
<tr>
<td></td>
<td>2:-</td>
<td>2:1</td>
<td>2:2</td>
</tr>
<tr>
<td>ve</td>
<td>1:2</td>
<td>1:1</td>
<td>1:-</td>
</tr>
<tr>
<td></td>
<td>2:-</td>
<td>2:-</td>
<td>2:-</td>
</tr>
<tr>
<td>la</td>
<td>1:-</td>
<td>1:-</td>
<td>1:-</td>
</tr>
<tr>
<td></td>
<td>2:-</td>
<td>2:2</td>
<td>2:1</td>
</tr>
</tbody>
</table>

based on this we get the following expected counts

\(^2\)to read this table the \((ve,gr)\) entry has 1:2 to indicate in first pair \((o^1, s^1)\), the \((ve,gr)\) pairing occurs in alignment \( a^1_2 \), and the pairing never occurs in the alignments for the second pair
and for these counts get new $tr(o|s)$ by normalising by column sums

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

You should complete two further iterations in the same fashion.