The Phrase-Based Translation Model

The Phrase Translation Model

- Simplest assumption is that all segmentations are equally likely, so there are no explicit probability terms for choice of segmentation.
- For any particular segmentations of \( o \) and \( s \) into \( K \) segments, the model of \( p(\bar{o}, \tau, \bar{s}) \) is:

\[
p(\bar{o}, \tau, \bar{s}) = p(\bar{o}, \tau | \bar{s}) \times p(\bar{s})
\]

\[
= \prod_{k=1}^{K} tr(\bar{o}_{\tau(k)} | \bar{s}_k) d(\bar{o}_{\tau(k-1)}, \bar{o}_{\tau(k)}) \]

\[
LM(s)
\]

- **Phrase translation**: there’s a term \( tr(\bar{o}_{\tau(k)} | \bar{s}_k) \) for the phrase-translation probabilities for an observed phrase \( \bar{o}_{\tau(k)} \) to be generated from source phrase \( \bar{s}_k \)
- **Reordering**: there is re-ordering term concerning how likely the destination for a \( \bar{s}_k \) phrase is given destination for previous \( \bar{s}_{k-1} \) phrase: \( d(\bar{o}_{\tau(k-1)}, \bar{o}_{\tau(k)}) \)
- **Language model**: the probability of the source phrases \( \bar{s} \) is equated to simply the probability of the source sequence \( s \) as given by an \( n \)-gram model \( LM(s) \)
The Phrase Translation Model contd

- using (2), preferred translation $s_{\text{best}}$ is defined as the source part of

$$\bar{s}, \tau = \arg \max_{\bar{s}, \tau} p(\bar{o}, \tau, \bar{s})$$

- note you could seek to define $s_{\text{best}}$ as

$$s_{\text{best}} = \arg \max_{\bar{s}} \sum_{\tau} p(\bar{o}, \tau, \bar{s})$$

but this is not standardly done; the above is regarded as a 'Viterbi' approximation of the sum

The distortion term

- The distortion term $d(\bar{o}_{r(k-1)}, \bar{o}_{r(k)})$ is not learned via EM
- instead just standardly defined as an exponentially decaying function of the 'distance' $x$ between end $\bar{o}_{r(k-1)}$ and start of $\bar{o}_{r(k)}$
- in particular where $x = |\text{fst}(\bar{o}_{r(k)}) - \text{lst}(\bar{o}_{r(k-1)})| - 1$, the $d$ term is $\alpha^x$ for some $\alpha < 1$.

illustration of Distance-Based Reordering

```
phrase | translated to | change $\bar{o}_{r(i-1)}$ to $\bar{o}_{r(i)}$ | ‘dislacement’ x
------ |-------------- |------------------------------- |---------------------
$\bar{s}_1$ | $\bar{o}_{r(1)} = o_1 : o_3$ | start at beginning | 0
$\bar{s}_2$ | $\bar{o}_{r(2)} = o_6 : o_6$ | $o_3$ to $o_6$ | $+2 = \text{fst}(o_6 : o_6) - \text{lst}(o_1 : o_3) - 1$
$\bar{s}_3$ | $\bar{o}_{r(3)} = o_4 : o_5$ | $o_6$ to $o_4$ | $-3 = \text{fst}(o_4 : o_5) - \text{lst}(o_6 : o_6) - 1$
$\bar{s}_4$ | $\bar{o}_{r(4)} = o_7 : o_7$ | $o_5$ to $o_7$ | $+1 = \text{fst}(\bar{o}_7 : \bar{o}_7) - \text{lst}(o_4 : o_5) - 1$
```

Scoring function: $\alpha^{|x|}$ — exponential with distance