Announcements
- A4 due Tues
- A5 out Tues, due the following Tues

Recap: Word alignment w/ IBM Model 1

\[ \bar{s} = \text{Je fais un bureau} \quad \text{NULL} \quad (s_1, \ldots, s_m, \text{NULL}) \]
\[ \bar{a} = 1 \quad [1, 2, 2, 3, 4] \]
\[ \bar{t} = \text{I am making a desk} \quad (t_1, \ldots, t_n) \]

\[ P(\bar{t}, \bar{a} | \bar{s}) = \prod_{i=1}^{n} P(a_i) P(t_i | s_{a_i}) \]

- Constant \( \frac{1}{m+1} \)
- Parameters \( P(t | s) \) matrix
- \( \prod_{s} \times |V_t| \)

Inference:

\[ P(a_i | \bar{s}, \bar{t}) \propto P(t_i | s_{a_i}) \]

Learning:

\[ \max_{\bar{a}} \sum_{j=1}^{D} \log P(\bar{t}(j), \bar{a}(j) | \bar{s}(j)) \quad \text{no labeled} \]

Instead:

\[ \max_{\bar{a}} \sum_{j=1}^{D} \log \sum_{\bar{a}} P(\bar{t}(j), \bar{a} | \bar{s}(j)) \quad \frac{P(\bar{t}(j) | \bar{s}(j))}{P(\bar{t}(j))} \]
Expectation Maximization: repeat E-M:
E-step: compute posteriors $P(a; i \in (i), \pi(i))$ on each train ex
  "labeling the date"
M-step: count + normalize using posteriors as labels (soft labels)

I do

Today: ① HMM alignment
       ② Phrase extraction
       ③ Phrase-based decoding
       ④ Syntactic MT
       ⑤ seq2seq models / NMT (next time)
HMM Alignment (Vogel, 1996)

Same as Model 1, except: \[ P(\bar{a}) = \prod_{i=1}^{n} P(\bar{a}_i | \bar{a}_{i-1}) \]

\[ a_i = 1 \ldots 20 \]

Phoneme Extraction

1. Run alignment model in both directions (Fr → En, En → Fr)
2. Intersect them
3. Find aligned phrases by drawing “boxes”
4. Count phrases, build phrase table
Qu’est-ce que tu fais

What are you doing

What are you doing

Qu’est-ce que tu fais

Requirements

- "Corners" of rectangle are aligned
- Words in phrase cannot align outside rectangle
Phrase-based decoding

One thing we could do:

\[
\max_T \sum_a P(a, T | \tilde{s}) \text{ using Model } 1 / HMM
\]

Word level translation doesn't work well

New modeling paradigm

Noisy channel (Shannon/Weaver 50s)

I see source \( \tilde{s} \)

Model is that there's an underlying "signal" \( \tilde{T} \), \( \tilde{s} \) is emitted from \( \tilde{T} \)

\[
P(\tilde{T} | \tilde{s}) \propto P(\tilde{s} | \tilde{T}) P(\tilde{T}) \rightarrow \text{ prior over } \tilde{T}
\]

likely \( \tilde{T} \), emission of \( \tilde{s} \)

View translation in "reverse" direction

TM is based on phrases