# CMPT 413 - Spring 2011 - Midterm #2

Please write down "Midterm #2" on the top of the answer booklet.

*When you have finished, return your answer booklet along with this question booklet.*

Mar 15, 2011

(1) Consider a language model over character sequences that computes the probability of a word based on the characters in that word, so if word $w = c_0, c_1, \ldots, c_n$ then $P(w) = P(c_0, \ldots, c_n)$. Let us assume that the language model is defined as a bigram character model $P(c_i \mid c_{i-1})$ where

$$P(c_0, \ldots, c_n) = \prod_{i=1,2,\ldots,n} P(c_i \mid c_{i-1}) \quad (1)$$

For convenience we assume that we have explicit word boundaries: $c_0 = \texttt{bos}$ and $c_n = \texttt{eos}$ where `bos` stands for *begin sentence marker* and `eos` stands for *end of sentence marker*.

Based on this model, for the English word *booking* the probability would be computed as:

$$P(booking) = P(b \mid \texttt{bos}) \times P(o \mid b) \times P(o \mid o) \times P(k \mid o) \times P(i \mid k) \times P(n \mid i) \times P(g \mid n) \times P(\texttt{eos} \mid g)$$

The inflection *ing* is a suffix and is generated after the stem *book* with probability

$$P(ing) = P(i \mid k) \times P(n \mid i) \times P(g \mid n) \times P(\texttt{eos} \mid g)$$

In Semitic languages, like Arabic and Hebrew, the process of inflection works a bit differently. In Arabic, for a word like *kitab* the stem would be *k-t-b* where the place-holders '-' for inflection characters have been added for convenience. We will assume that each word is made up of a sequence of consonant-vowel sequences CVCVCV... and the vowels always form the inflection.

a. Provide the definition of an $n$-gram model that will compute the probability for the word *kitab* and *k-t-b* as follows:

$$P(kitab) = P(k \mid \texttt{bos}) \times P(t \mid k) \times P(b \mid t) \times P(i \mid b) \times P(a \mid i) \times P(\texttt{eos} \mid a)$$

$$P(k\text{-}t\text{-}b) = P(k \mid \texttt{bos}) \times P(t \mid k) \times P(b \mid t) \times P(\text{-} \mid b) \times P(\text{-} \mid \text{-}) \times P(\texttt{eos} \mid \text{-})$$

Write down the equation for this $n$-gram model in the same mathematical notation as equation (1).

*Answer:*

$$P(c_0, \ldots, c_n) = \begin{cases} \prod_{i=1}^{n} P(c_i \mid c_{i-1}) & \text{if } n \leq 3 \\ \Big(P(c_1 \mid c_0) \times \prod_{i=3,5,\ldots}^{\ell} P(c_i \mid c_{i-2})\Big) \times \\ \Big(P(c_2 \mid c_{\ell_o}) \times \prod_{i=4,6,\ldots}^{\ell} P(c_i \mid c_{i-2}) \times P(c_n \mid c_{\ell_e})\Big) & \text{if } n > 3 \end{cases}$$

Define $\ell = n - (n \mod 2)$ and $\ell_o$ is the last odd number less than $\ell$ and $\ell_e$ is the last even number less than $\ell$. As long as the boundary cases are right for the bigrams, we don't penalize off by one in the length, and we don't penalize for $n \leq 3$.

b. Using your $n$-gram model show how $P(kitab) = P(ktb) \times P(ia)$.

*Answer:*

$$\begin{aligned} P(kitab) &= P(c_0 = \texttt{bos}, c_1 = k, c_2 = i, c_3 = t, c_4 = a, c_5 = b, c_6 = \texttt{eos}) \\ &= P(ktb) \times P(ia, \texttt{eos}) \\ P(ktb) &= P(c_1 = k \mid c_0 = \texttt{bos}) \times P(c_3 = t \mid c_1 = k) \times P(c_5 = b \mid c_3 = t) \\ &\quad \text{this term corresponds to the first bracket in the eqn above} \\ P(ia) &= P(c_2 = i \mid c_{\ell_o} = c_5 = b) \times P(c_4 = a \mid c_2 = i) \times P(c_n = c_6 = \texttt{eos} \mid c_{\ell_e} = c_4 = a) \\ &\quad \text{corresponds to the second bracket in the eqn above} \end{aligned}$$

(2) The probability model $P(t_i \mid t_{i-2}, t_{i-1})$ is provided below where each $t_i$ is a part of speech tag, e.g. $P(\mathrm{D} \mid \mathrm{N}, \mathrm{V}) = \frac{1}{3}$. Also provided is $P(w_i \mid t_i)$ that a word $w_i$ has a part of speech tag $t_i$, e.g. $P(\text{flies} \mid \mathrm{V}) = \frac{1}{2}$.
The part of speech tag definitions are: `bos` (*begin sentence marker*), N (*noun*), V (*verb*), D (*determiner*), P (*preposition*), `eos` (*end of sentence marker*).

| $P(t_i \mid t_{i-2}, t_{i-1})$ | $t_{i-2}$ | $t_{i-1}$ | $t_i$ |
|---|---|---|---|
| 1 | bos | bos | N |
| $\frac{1}{2}$ | bos | N | N |
| $\frac{1}{2}$ | bos | N | V |
| $\frac{1}{2}$ | N | N | V |
| $\frac{1}{2}$ | N | N | P |
| $\frac{1}{3}$ | N | V | D |
| $\frac{1}{3}$ | N | V | V |
| $\frac{1}{3}$ | N | V | P |
| 1 | V | D | N |
| 1 | V | V | D |
| 1 | N | P | D |
| 1 | V | P | D |
| 1 | P | D | N |
| 1 | D | N | eos |

| $P(w_i \mid t_i)$ | $t_i$ | $w_i$ |
|---|---|---|
| 1 | D | an |
| $\frac{2}{5}$ | N | time |
| $\frac{2}{5}$ | N | arrow |
| $\frac{1}{5}$ | N | flies |
| 1 | P | like |
| $\frac{1}{2}$ | V | like |
| $\frac{1}{2}$ | V | flies |
| 1 | eos | eos |
| 1 | bos | bos |

a. Consider a Jelinek-Mercer style interpolation smoothing scheme for $P(w_i \mid t_i)$:

$$P_{jm}(w_i \mid t_i) = \Lambda[t_i] \cdot P(w_i \mid t_i) + (1 - \Lambda[t_i]) \cdot P(w_i)$$

$\Lambda$ is an array with a value $\Lambda[t_i]$ for each part of speech tag $t_i$, such that $0 \leq \Lambda[t_i] \leq 1$. Provide a condition on $\Lambda$ that must be satisfied to ensure that $P_{jm}$ is a well-defined probability model.

*Answer:* Because of the following fact about $P(w_i \mid t_i)$:

$$\sum_{w_i} P(w_i \mid t_i) = 1$$

and in $P_{jm}$ we are given $t_i$, so to interpolate with $P(w_i)$ the following condition has to hold:

$$\sum_{t_i} \Lambda[t_i] = 1$$

b. Provide a Hidden Markov Model (*hmm*) that uses the trigram part of speech probability $P(t_i \mid t_{i-2}, t_{i-1})$ as the transition probability $P_{hmm}(s_j \mid s_k)$ and the probability $P(w_i \mid t_i)$ as the emission probability $P_{hmm}(w_j \mid s_j)$.

**Important:** Provide the *hmm* in the form of two tables as shown below. The first table contains transitions between states in the *hmm* and the transition probabilities and the second table contains the words emitted at each state and the emission probabilities. Do not provide entries with zero probability.

| from-state $s_k$ | to-state $s_j$ | $P(s_j \mid s_k)$ |
|---|---|---|
| | | |

| state $s_j$ | emission $w$ | $P(w \mid s_j)$ |
|---|---|---|
| | | |

*Hint:* In your *hmm* the state $\langle N, \texttt{eos} \rangle$ will have emission of word `eos` with probability 1 and will not have transitions to any other states.

*Answer:* Here are the two tables that define the HMM, the transition table on the left and the emission table on the right:

| from-state $s_k$ | to-state $s_j$ | $P(s_j \mid s_k)$ | |
|---|---|---|---|
| $bos, bos$ | $bos, N$ | $P(N \mid bos, bos)$ | 1 |
| $bos, N$ | $N, N$ | $P(N \mid bos, N)$ | $\frac{1}{2}$ |
| $bos, N$ | $N, V$ | $P(V \mid bos, N)$ | $\frac{1}{2}$ |
| $N, N$ | $N, V$ | $P(V \mid N, N)$ | $\frac{1}{2}$ |
| $N, N$ | $N, P$ | $P(P \mid N, N)$ | $\frac{1}{2}$ |
| $N, V$ | $V, D$ | $P(D \mid N, V)$ | $\frac{1}{3}$ |
| $N, V$ | $V, V$ | $P(V \mid N, V)$ | $\frac{1}{3}$ |
| $N, V$ | $V, P$ | $P(P \mid N, V)$ | $\frac{1}{3}$ |
| $V, D$ | $D, N$ | $P(N \mid V, D)$ | 1 |
| $V, V$ | $V, D$ | $P(D \mid V, V)$ | 1 |
| $N, P$ | $P, D$ | $P(D \mid N, P)$ | 1 |
| $V, P$ | $P, D$ | $P(D \mid V, P)$ | 1 |
| $P, D$ | $D, N$ | $P(N \mid P, D)$ | 1 |
| $D, N$ | $N, eos$ | $P(eos \mid D, N)$ | 1 |

| state $s_j$ | emission $w$ | $P(w \mid s_j)$ |
|---|---|---|
| $bos, bos$ | $bos$ | 1 |
| $bos, N$ | time | $\frac{2}{5}$ |
| $bos, N$ | arrow | $\frac{2}{5}$ |
| $bos, N$ | flies | $\frac{1}{5}$ |
| $N, N$ | time | $\frac{2}{5}$ |
| $N, N$ | arrow | $\frac{2}{5}$ |
| $N, N$ | flies | $\frac{1}{5}$ |
| $N, V$ | like | $\frac{1}{2}$ |
| $N, V$ | flies | $\frac{1}{2}$ |
| $V, D$ | an | 1 |
| $V, V$ | like | $\frac{1}{2}$ |
| $V, V$ | flies | $\frac{1}{2}$ |
| $N, P$ | like | 1 |
| $V, P$ | like | 1 |
| $P, D$ | an | 1 |
| $D, N$ | time | $\frac{2}{5}$ |
| $D, N$ | arrow | $\frac{2}{5}$ |
| $D, N$ | flies | $\frac{1}{5}$ |

c. Based on your *hmm* constructed in 2b. what is the state sequence that would be provided by the Viterbi algorithm for the following input sentence:

`bos bos` time flies like an arrow `eos`

*Answer:*
Note that the only ambiguous words are *flies* (could be $N$ or $V$) and *like* (could be $V$ or $P$) and so all you need to do is compare the scores for the following sub-sequence. The bold-faced outcome wins for this sub-sequence which determines the best state sequence for the entire input.

| flies | like | |
|---|---|---|
| **(N, V)** | **(V, P)** | $\mathbf{\frac{1}{2} \times \frac{1}{3} \times 1}$ |
| $(N, V)$ | $(V, V)$ | $\frac{1}{2} \times \frac{1}{3} \times \frac{1}{2}$ |
| $(N, N)$ | $(N, P)$ | $\frac{1}{5} \times \frac{1}{2} \times 1$ |
| $(N, N)$ | $(N, V)$ | $\frac{1}{5} \times \frac{1}{2} \times \frac{1}{2}$ |

Since the best state sequence is then (bos,N)–(N,V)–(V,P)–(P,D)–(D,N)–(N,eos) the output best state sequence will be *bos/bos, time/N, flies/V, like/P, an/D, arrow/N, eos/eos*.

*Answer:* The full table is given below but you do not need to compute the entire table to solve this question.

| bos | time | flies | like | an | arrow | eos | |
|---|---|---|---|---|---|---|---|
| (bos,bos) 1 | (bos,N) $\times 1 \times \frac{2}{5}$ | (N,V) $\times \frac{1}{2} \times \frac{1}{2}$ | (V,P) $\times \frac{1}{3} \times 1$ | (P,D) $\times 1 \times 1$ | (D,N) $\times 1 \times \frac{2}{5}$ | (N,eos) $\times 1 \times 1$ | $= \frac{1}{75}$ |
| (bos,bos) 1 | (bos,N) $\times 1 \times \frac{2}{5}$ | (N,V) $\times \frac{1}{2} \times \frac{1}{2}$ | (V,V) $\times \frac{1}{3} \times \frac{1}{2}$ | (V,D) $\times 1 \times 1$ | (D,N) $\times 1 \times \frac{2}{5}$ | (N,eos) $\times 1 \times 1$ | $= \frac{1}{150}$ |
| (bos,bos) 1 | (bos,N) $\times 1 \times \frac{2}{5}$ | (N,N) $\times \frac{1}{2} \times \frac{1}{5}$ | (N,P) $\times \frac{1}{2} \times 1$ | (P,D) $\times 1 \times 1$ | (D,N) $\times 1 \times \frac{2}{5}$ | (N,eos) $\times 1 \times 1$ | $= \frac{1}{125}$ |
| (bos,bos) 1 | (bos,N) $\times 1 \times \frac{2}{5}$ | (N,N) $\times \frac{1}{2} \times \frac{1}{5}$ | (N,V) $\times \frac{1}{2} \times \frac{1}{2}$ | (V,D) $\times \frac{1}{3} \times 1$ | (D,N) $\times 1 \times \frac{2}{5}$ | (N,eos) $\times 1 \times 1$ | $= \frac{1}{750}$ |