Homework 2

in notes

N1.1

Note that I've used M. Collins notation presented on his website cs.columbia.edu/~mcollins

Transition probabilities

start/end symbol

$q(N|\$) = 0.5 \quad q(N|N) = 0.3 \quad q(\$|V) = 0.5 \quad q(A|A) = 0.4$

$q(A|\$) = 0.5 \quad q(V|N) = 0.5 \quad q(N|V) = 0.5 \quad q(N|A) = 0.6$

$q(\$|N) = 0.2$

Viterbi excerpt

...

For $k = 1..n$

For $v \in K_k$

$\pi(k,v) = \max_{w \in K_{k-1}} (\pi(k-1,w) \times q(v|w) \times e(x_k|v))$

$bp(k,v) = \arg\max_{w \in K_{k-1}} \pi(k-1,w) \times q(v|w) \times e(x_k|v)$

...

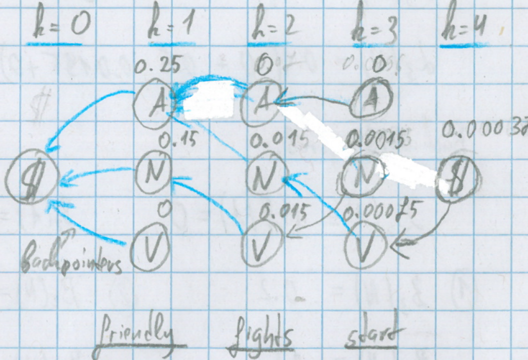
Emission probabilities

$e(friendly|A) = 0.5 \quad e(fights|A) = 0 \quad e(start|A) = 0.7$

$e(friendly|N) = 0.3 \quad e(fights|N) = 0.1 \quad e(start|N) = 0.2$

$e(friendly|V) = 0.2 \quad e(fights|V) = 0.2 \quad e(start|V) = 0.1$

Calculations



① $\pi(1,A) = \pi(0,\$) \cdot q(A|\$) e(friendly|A) = 1 \times 0.5 \times 0.5 = 0.25$

$\pi(1,N) = 1 \times 0.5 \times 0.3 = 0.15$

$\pi(1,V) = 1 \times 0 \times 0.2 = 0$

② $\pi(2,A) = \max(\pi(1,A) \times q(A|A) \times e(fight|A),$
$\pi(1,N) \times q(A|N) \times e(fight|A),$
$\pi(1,V) \times q(A|V) \times e(fight|A)) = \max(0, 0, 0) = 0$

$\pi(2,N) = \max((0.25 \times 0.6 \times 0.1), (0.15 \times 0.3 \times 0.1), 0) = \max(0.015, 0.0045, 0) = 0.015$

$\pi(2,V) = \max(0, 0.15 \times 0.5 \times 0.2, 0) = 0.015$

③ $\pi(3,A) = \max(0,0,0) = 0$

$\pi(3,N) = \max(0.0 \times 0.6 \times 0.2, 0.015 \times 0.3 \times 0.2, 0.015 \times 0.5 \times 0.2) = \max(0.0, 0.0009, \overset{\downarrow}{\overbrace{0.0015}}) = 0.0015$

$\pi(3,V) = \max(0.0 \times 0 \times 0.1, 0.015 \times 0.5 \times 0.1, 0.015 \times 0 \times 0.1) = 0.00075$

④ $y_3 = \arg\max_{v \in K_n} (\pi(n,v) \times q(\$|v)) = \arg\max_{v} (0.196 \times 0, 0.0015 \times 0.2, 0.00075 \times 0.5) =$

$= \arg\max_{v} (0, 0.0003, 0.000375) = V$

$y_2 = N$ ← obtained from back pointers

$y_1 = A$

Probability of the sequence $\langle A, N, V \rangle$ is: 0.000375

N1.2

Forward prob.

$\alpha_1(A) = p(x_1, S_1 = A) = q(A|\varnothing)\, e(\text{friendly}|A) = 0.5 \times 0.5 = 0.25$

(1) $\alpha_1(N) = q(N|\varnothing)\, e(\text{friendly}|N) = 0.15$

$\alpha_1(V) = 0$

set of states

$\alpha_2(A) = \sum_{j=1}^{3} e(\text{lights}|A)\, p(A|S_j)\, \alpha_1(j) = 0(0.4 \cdot 0.25 + 0 + 0) = 0$

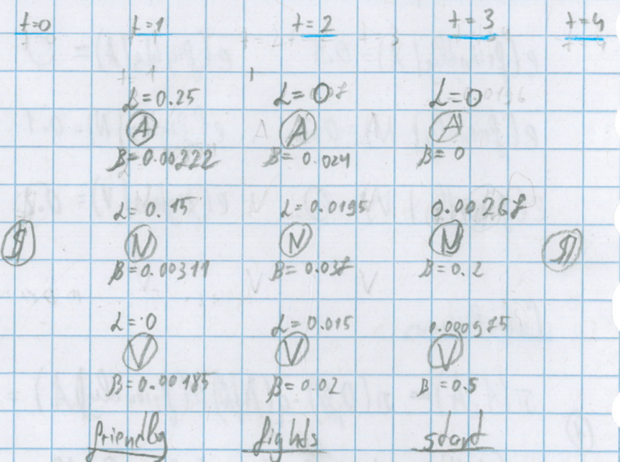
(2) $\alpha_2(N) = 0.1(0.15 + 0.045 + 0) = 0.0195$

$\alpha_2(V) = 0.2(0 + 0.5 \cdot 0.15 + 0) = 0.015$

(3) $\alpha_3(A) = 0.7(0.4 \times 0.0 + 0 + 0) = 0$

$\alpha_3(N) = 0.2(0.0 + 0.00585 + 0.0075) = 0.00267$

$\alpha_3(V) = 0.1(0 + 0.5 \cdot 0.0195 + 0) = 0.000975$



Backward prob.

$\beta_3(A) = q(\varnothing|A) = 0$ $\qquad \beta_2(A) = \sum_{j=1}^{M} \beta_3(j)\, q(S_j|A)\, e(x_3|S_j) = 0 + 0.2 \cdot 0.6 \cdot 0.2 + 0.5 \cdot 0 = 0.024$

(1) $\beta_3(N) = 0.2$ $\qquad$ (2) $\beta_2(N) = 0 + 0.2 \cdot 0.3 \cdot 0.2 + 0.5 \cdot 0.5 \cdot 0.1 = 0.012 + 0.025 = 0.037$

$\beta_3(V) = 0.5$ $\qquad \beta_2(V) = 0 + 0.2 \cdot 0.5 \cdot 0.2 + 0 = 0.02$

(3) $\beta_1(A) = 0.024 \cdot 0.4 \cdot 0 + 0.037 \cdot 0.6 \cdot 0.1 + 0.02 \cdot 0 = 0.00222$

$\beta_1(N) = 0.024 \cdot 0 + 0.037 \cdot 0.3 \cdot 0.1 + 0.02 \cdot 0.5 \cdot 0.2 = 0.00111 + 0.002 = 0.00311$

$\beta_1(V) = 0.024 \cdot 0 + 0.037 \cdot 0.5 \cdot 0.1 + 0 = 0.00185$

Cheap way to test the correctness of computations: $P(O) = \sum_{i=1}^{M} \alpha_n(i)\, q(\varnothing|i) = \sum_{i=1}^{M} q(i|\varnothing)\, e(o_1|i)\, \beta_1(i)$

α: $0 + 0.00267 \times 0.2 + 0.000975 \cdot 0.5 = 0.0010215$

β: $0.00222 \times 0.5 \cdot 0.5 + 0.00311 \cdot 0.5 \cdot 0.3 + 0 = 0.0010215$

N1.3

1) $p(S_t = j | x_1, \dots, x_n) = \dfrac{p(x_{1:n}, S_t = j)}{p(x_{1:n})} = \dfrac{p(x_{1:t}, x_{t+1:n}, S_t = j)}{\sum_{j=1}^{M} p(x_{1:t}, x_{t+1:n}, S_t = j)} = \dfrac{p(x_{1:t}, S_t = j)\, p(x_{t+1:n}|S_t = j)}{\sum_{j'=1} p(x_{1:t}, S_t = j')\, p(x_{t+1:n}|S_t = j')}$

$= \dfrac{\alpha_t(j)\, \beta_t(j)}{\sum_{j'=1}^{M} \alpha_t(j')\, \beta_t(j')}$ ← formal explanation of how forward & backward prob. are used.

1) $p(S_3 = N | x_{1:n}) = \dfrac{0.00267 \times 0.2}{0 + 0.00267 \times 0.2 + 0.000975 \cdot 0.5} = \dfrac{0.000534}{0.0010215} = 0.52276065$

MAP will choose N.

N 1.3 (contined)

2) $p(S_3 = V \mid x_{1:n}) = \frac{0.0004875}{0.0010215} = 0.47723535$

3) $p(S_3 = A \mid x_{1:n}) = 0$

N 1.4

Viterbi bases decisions of next states on maximum likelihood estimate and considers only one state that has maximum probability. It completely ignores the fact that there might be many states that point to the state N to be the next one.

On the other hand MAP takes into account all possible paths and not only only the most likely ~~maximum likelihood to be the proper one~~

It's easy to see why MAP indicates that t=3 tag is N based on transitions. I.e. we have transitions to the tag N from A, V, N while only one a transition probability from N to V. Note that, „start" has very similar emission from both tags, so choice based on emission is not affected much.

N 1.5

In case of Viterbi we decode (select) the most likeli sequence of tags, ie a path. While in MAP we maximize the probability of a concrete tag given observations, by considering all paths that lead to the tag.

Therefore, if we want to have max per word accuracy we should choose MAP, and Viterbi if we want to predict the whole sequence correctly.

N2

(1,1) $\beta_{NP}(1,1) = p(NP \to N)\, p(N \to friendly) = 0.2 \times 0.3 = 0.06$

$\beta_S(1,1) = p(S \to NP) \times \beta_{NP}(1,1) = 0.5 \times 0.06 = 0.03$

(2,2) $\beta_{NP}(2,2) = p(NP \to N) \times p(N \to fights) = 0.2 \times 0.1 = 0.02$

$\beta_{VP}(2,2) = p(VP \to V) \times p(V \to fights) = 0.6 \times 0.2 = 0.12$

$\beta_S(2,2) = p(S \to NP) \times \beta_{NP}(2,2) = 0.5 \times 0.02 = 0.01$

(3,3) $\beta_{NP}(3,3) = p(NP \to N)\, p(N \to start) = 0.2 \times 0.2 = 0.04$

$\beta_{VP}(3,3) = p(VP \to V) \times p(V \to start) = 0.6 \times 0.1 = 0.06$

$\beta_S(3,3) = p(S \to NP) \times \beta_{NP}(3,3) = 0.5 \times 0.04 = 0.02$

(1,2) $\beta_{NP}(1,2) = p(NP \to N\ NP) \times \beta_N(1,1) \times \beta_{NP}(2,2) + p(NP \to A\ NP) \times \beta_A(1,1) \times \beta_{NP}(2,2) = 0.5 \times 0.3 \times 0.02 + 0.3 \times 0.5 \times 0.02$
$= 0.006$

$\beta_S(1,2) = p(S \to NP\ VP) \times \beta_{NP}(1,1) \times \beta_{VP}(2,2) = 0.5 \times 0.06 \times 0.12 = 0.0036$

(2,3) $\beta_{NP}(2,3) = p(NP \to N\ NP) \times \beta_N(2,2) \times \beta_{NP}(3,3) = 0.5 \times 0.1 \times 0.04 = 0.002$

$\beta_{VP}(2,3) = p(VP \to V\ NP) \times \beta_V(2,2) \times \beta_{NP}(3,3) = 0.4 \times 0.2 \times 0.04 = 0.0032$

$\beta_S(2,3) = p(S \to NP\ VP) \times \beta_{NP}(2,2) \times \beta_{VP}(3,3) = 0.5 \times 0.02 \times 0.06 = 0.0006$

(1,3) $\beta_S(1,3) = p(S \to NP\ VP)\, \beta_{NP}(1,1) \times \beta_{VP}(2,3) + p(S \to NP\ VP) \times \beta_{NP}(1,2) \times \beta_{VP}(3,3) +$

$+\ p(S \to NP)\ \beta_{NP}(1,3) = 0.5 \times 0.06 \times 0.0032 + 0.5 \times 0.006 \times 0.06 + 0.5 \times 0.0006 = 0.000576$

	1	2	3
1	A: 0.5 N: 0.3 NP: 0.06 S: 0.03	NP: 0.006 S: 0.0036	S: 0.000576
2		N: 0.1 V: 0.2 NP: 0.02 VP: 0.12 S: 0.01	NP: 0.002 VP: 0.0032 S: 0.0006
3			N: 0.2 V: 0.1 NP: 0.04 VP: 0.06 S: 0.02

$P(\text{friendly}, \text{fights}, \text{start} \mid G) = \beta_S(1,3) = 0.000576$

N3

Inside prob.

(1,1) $\beta_N(1,1) = 0.1$

(2,3) 1) $\beta_{VP[1]}(2,3) = p(VP[1] \to M\ V[1])\, \beta_M(2,2)\, \beta_{V[1]}(3,3) + p(VP[1] \to M\ V[2])$

$\times\ \beta_M(2,2)\ \beta_{V[2]}(3,3) = 0.00002 + 0.0002 = 0.00022$

(2,3) 2) $\beta_{VP[2]}(2,3) = p(VP[2] \to M\ V[1]) \times \beta_M(2,2)\, \beta_{V[1]}(3,3) + p(VP[2] \to M\ V[2])\, \beta_M(2,2)\, \beta_{V[2]}(3,3) =$

$= 0.00004 + 0.04 = 0.00404$

(1,3) $\beta_S(1,3) = 0.3(0.01 \times 0.00022 + 0) + 0.2(0.01 \times 0.00404 + 0) = 0.000,008,74$

Outside prob

(1,3) $\alpha_S(1,3) = 1$

$\alpha_{VP[1]}(2,3) = \alpha_S(1,3) \times \beta_{NP}(1,1) \times p(S \to NP\ VP[1]) = 1 \times 0.01 \times 0.3 = 0.003$

$\alpha_{VP[2]}(2,3) = \alpha_S(1,3) \times \beta_{NP}(1,1) \times p(S \to NP\ VP[2]) = 0.002$

(3,3) $\alpha_{V[1]}(3,3) = \alpha_{VP[1]}(2,3)\, \beta_M(2,2)\, p(VP[1] \to M\ V[1]) + 0 + \alpha_{VP[2]}(2,3)\, \beta_M(2,2) \times p(VP[2] \to M\ V[1]) =$

$= 0.000,006 + 0.000,008 = 0.000,014$

$\alpha_{V[2]}(3,3) = 0.003 \times 0.2 \times 0.01 + 0.002 \times 0.2 \times 0.2 = 0.000,086$

N 3 (continued)

Expected counts (E-step):

1) $E[C(S \to NP\ VP[1])|T] = \dfrac{\sum_{p=1}^{3}\sum_{q=p+1}^{3} \alpha_S(p,q)\, P(S \to NP\ VP[1]) \sum_{d=p}^{2} \beta_{NP}(p,d)\, \beta_{VP[1]}(d+1,q)}{\beta_S(1,3)} =$

$= \alpha_S(1,3) \times P(S \to NP\ VP[1]) \left(\beta_{NP}(1,1) \times \beta_{VP[1]}(2,3)\right) \cdot \dfrac{1}{\beta_S(1,3)} = \dfrac{0.000,000,66}{0.000,008,74} = 0.07551487$

2) $E[C(S \to NP\ VP[2])|T) = \alpha_S(1,3) \times P(S \to NP\ VP[2]) \times \beta_{NP}(1,1) \times \beta_{VP[2]}(2,3) \cdot \dfrac{1}{\beta_S(1,3)} = 0.92448513$

3) $E[C(VP[1] \to M\ V[1])|T] = \alpha_{VP[1]}(2,3) \times P(VP[1] \to M\ V[1]) \times \beta_M(2,2) \times \beta_{V[1]}(3,3) \cdot \dfrac{1}{\beta_S(1,3)} =$

$= \dfrac{0.000,000,06}{\beta_S(1,3)} = 0.006,864,99$

4) $E[C(VP[1] \to M\ V[2])|T] = 0.003 \times 0.01 \times 0.2 \times 0.1 \cdot \dfrac{1}{\beta_S(1,3)} = 0.068,6499$

5) $E[C(VP[2] \to M\ V[1])|T] = 0.002 \times 0.02 \times 0.2 \times 0.01 \cdot \dfrac{1}{\beta_S(1,3)} = 0.009,153,32$

6) $E[C(VP[2] \to M\ V[2])|T] = 0.002 \times 0.2 \times 0.2 \times 0.1 \cdot \dfrac{1}{\beta_S(1,3)} = 0.91533181$

7) $E[C(V[1] \to \text{poison})|T] = 0.000,014 \times 0.01 \cdot \dfrac{1}{\beta_S(1,3)} = 0.01601831$

8) $E[C(V[2] \to \text{poison})|T] = 0.000,086 \times 0.1 \cdot \dfrac{1}{\beta_S(1,3)} = 0.98398165$

Probabilities reestimation (M-step):

1) $P(S \to NP\ VP[1]|T) = \dfrac{E[C(S \to NP\ VP[1])|T]}{E[C(S \to NP\ VP[1])|T] + E[C(S \to NP\ VP[2])|T]} =$

$= \dfrac{0.07551487}{1} = 0.07551487$

2) $P(S \to NP\ VP[2]|T) = 0.92448513$

3) $P(VP[1] \to M\ V[1]|T) = \dfrac{0.00686499}{0.07551489} \approx 0.091$

4) $P(VP[1] \to M\ V[2]|T) \approx 0.909$

5) $P(VP[2] \to M\ V[1]|T) = \dfrac{0.00915332}{0.92448513} \approx 0.01$

6) $P(VP[2] \to M\ V[2]|T) \approx 0.99$

7) $P(V[1] \to \text{poison}|T) = P(V[2] \to \text{poison}|T) = 1$ (either it's a mistake in the task or those rules do have the same probs.)