Damianos Dimitrios
03119825

(1)

1) $y(n) = v(n) * h_2(n)$ με $v(n) = x(n) * h_1(n)$, άρα $y(n) = (x(n) * h_1(n)) * h_2(n) \Rightarrow$

$$\Rightarrow y(n) = \left(\sum_{k=-\infty}^{+\infty} x(k) h_1(n-k)\right) * h_2(n) = \left(\sum_{k=-\infty}^{+\infty} \sum_{\gamma=-\infty}^{+\infty} x(k) h_1(n-k) h_2(n-k-\gamma)\right)$$

$$\Rightarrow y(n) = \sum_{k=-\infty}^{+\infty} x(k) \left(\sum_{\gamma=-\infty}^{+\infty} h_1(n-k) \cdot h_2(n-k-\gamma)\right) = \sum_{k=-\infty}^{+\infty} x(k) \cdot h(n-k)$$

με $h(n) = \sum_{\gamma=-\infty}^{+\infty} h_1(\gamma) h_2(n-\gamma)$. Άρα $y(n) = x(n) * h(n)$, με $\boxed{h(n) = h_1(n) * h_2(n)}$

Τελικά η κρουστική απόκριση του συνολικού συστήματος είναι $h(n) = h_1(n) * h_2(n)$

2) νδο $h_1(n) * h_2(n) = h_2(n) * h_1(n)$

$$h_1(n) * h_2(n) = \sum_{k=-\infty}^{+\infty} h_1(k) \cdot h_2(n-k) \xrightarrow[k=\gamma-n]{\gamma=n-k} \sum_{\gamma-n=-\infty}^{+\infty} h_1(\gamma-n) h_2(\gamma) = \sum_{\gamma=-\infty}^{+\infty} h_2(\gamma) \cdot h_1(\gamma-n) =$$

n = σταθερό κάθε φορά

$$= h_2(n) * h_1(n)$$

3) $$H(z) = \left(\sum_{r=0}^{M} b_r \cdot z^{-r}\right)\left(\frac{1}{1 - \sum_{k=1}^{N} a_k z^{-k}}\right) = H_1(z) \cdot H_2(z)$$

$$Y(z) = H(z) X(z) \Rightarrow \left(1 - \sum_{k=1}^{N} a_k z^{-k}\right) Y(z) = \left(\sum_{r=0}^{M} b_r z^{-r}\right) X(z) \Rightarrow$$

$$\Rightarrow Y(z) - \sum_{k=1}^{N} a_k z^{-k} Y(z) = \sum_{r=0}^{M} b_r z^{-r} X(z) \xrightarrow{Z^{-1}(\cdot)}$$

$$\Rightarrow y(n) - \sum_{k=1}^{N} a_k y(n-k) = \sum_{r=0}^{M} x(n-r)$$ η εξίσωση διαφορών

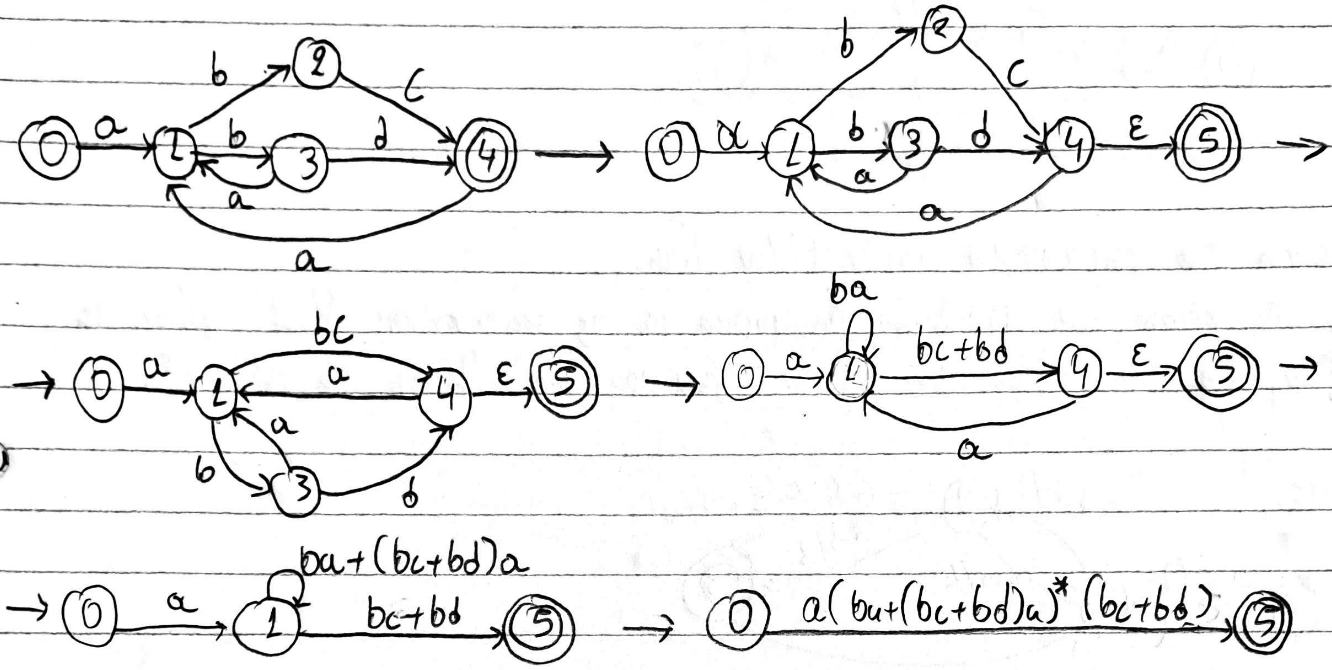
4) $H(z) = H_2(z) \cdot H_1(z)$

$Y(z) = H(z) \cdot X(z) = H_2(z) \cdot H_1(z) \cdot X(z) \Rightarrow \frac{Y(z)}{H_2(z)} = H_1(z) \cdot X(z)$, καταλήγουμε στο ίδιο

Η σειρά δεν παίζει ρόλο

(2)

1) Για την εύρεση της κανονικής μορφής εργαζόμαστε ως εξής:
Αφαιρούμε διαδοχικά ενδιάμεσες καταστάσεις, ενημερώνοντας κατάλληλα τις μεταβάσεις.



Άρα η κανονική έκφραση είναι $a(ba+(bc+bd)a)^*(bc+bd)$

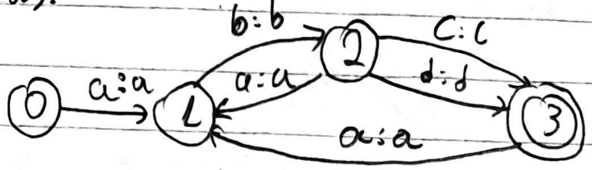
2) Η πιο πιθανή γραμματοσειρά που αποδέχεται, εφόσον χρησιμοποιεί τον τροπικό ημιδακτύλιο, είναι αυτή με το μικρότερο κόστος (άθροισμα του min του ημιδακτυλίου)
Συνεπώς, 2 είναι οι υποψήφιες (και μικρότερες συμβολοσειρές) που το πετυχαίνουν:
"abc" και "abd"
$cost(\text{"abc"}) = 2 \otimes 1,5 \otimes 2,5 = 2+1,5+2,5 = 6$
$cost(\text{"abd"}) = 2 \otimes 1 \otimes 3 = 2+1+3 = 6$
Άρα και οι δύο είναι πιθανές

Να σημειωθεί ότι τα μονοπάτια που ακολουθούν είναι:

"abc": $0 \to 1 \to 2 \to 4$
"abd": $0 \to 1 \to 3 \to 4$

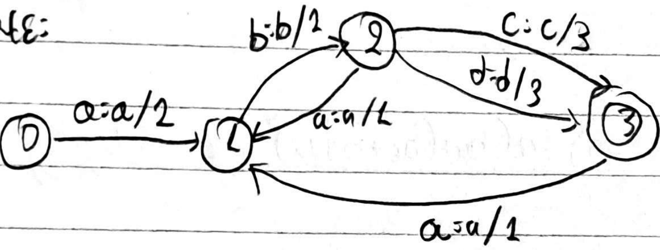
4) Η ντετερμινιστική μηχανή προκύπτει αν ενώσουμε τις 2, 3, καθώς δέχονται τις ίδιες εισόδους.

Άρα



5) Τα κόστη των μεταβάσεων αυτών είναι ίδια.
Όμως τα κόστη των μεταβάσεων αναφορικά με τις καταστάσεις 2,3 πρέπει να αλλάξουν, ώστε το τελικό κόστος των λέξεων που αποδέχεται να είναι ίδιο.

Άρα έχουμε:



Παρατηρούμε ότι $1 \xrightarrow{b} 2$ άρα κόστος 1, ενώ στο αρχικό FSM έχουμε
b:b/1 και b:b/1,5
Το επιπλέον 0,5 που "χάνουμε" από το bc μπορούμε να το προσθέσουμε στο c
Άρα προκύπτει c:c/2,5 → c:c/3

3) Για τη συμβολοσειρά "abdababc" εξετάζουμε το αρχικό FSM και έχουμε τις εξής μεταβάσεις καταστάσεων:

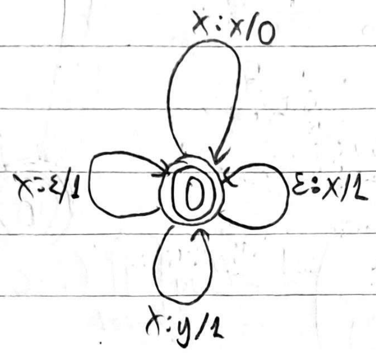
$$0 \xrightarrow{a} 1 \xrightarrow{b} 3 \xrightarrow{d} 4 \xrightarrow{a} 1 \xrightarrow{b} 3 \xrightarrow{a} 1 \xrightarrow{b} 2 \xrightarrow{c} 4$$

Το παραπάνω μονοπάτι είναι το μοναδικό ώστε το FSM να αποδεχτεί την συμβολοσειρά, και το κόστος είναι:

$$2 + 1 + 3 + 1 + 1 + 1 + 1,5 + 2,5 = \underline{13}$$

(3) $\Sigma = \{A, G, C, T, E, F\}$

1) Η μορφή του Levenshtein transducer είναι:



για $x \neq y$ και
για κάθε $x, y \in \Sigma$ έχουμε ένα μονοπάτι

Για λόγους απλότητας, δίνουμε αυτή τη μορφή.

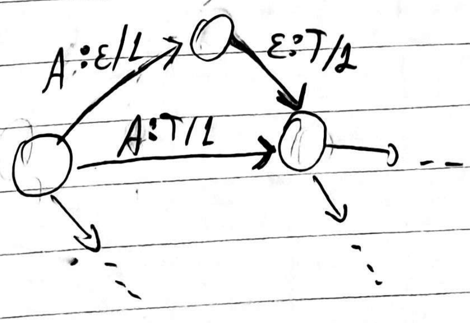
2) Για να βρούμε την φθηνότερη αντιστοίχιση, δημιουργούμε δύο αποδοχείς, ο καθένας αποδέχεται μία από τις ακολουθίες, έστω acc1 (για τη πρώτη) και acc2 (για τη 2η)
Συνθέτοντας το acc1 με το Levenshtein transducer, προκύπτει ένα FST που μας δίνει όλες τις ακολουθίες που προκύπτουν από την 1η ακολουθία και τις μετατροπές που ορίζει ο Levenshtein.
Τέλος, συνθέτουμε το παραγόμενο FST με το acc2, ώστε να πάρουμε ένα FST που περιέχει όλες τις μετατροπές που οδηγούν από την 1η στην 2η ακολουθία.
Βρίσκοντας το shortest path στο τελευταίο FST, βρίσκουμε το ελάχιστο κόστος.

Το ελάχιστο κόστος είναι 5:

A E C A G E F
1↓ 0↓ 1↓ 1↓ 0↓ 1↓ 1↓
T E T C G A G

κάνουμε αντικατάσταση των χαρακτήρων που δεν ταιριάζουν

3) Για τη 2η καλύτερη, θα έχουμε μετατροπή κάποιου χαρακτήρα σε ε, και μετά από ε στον χαρακτήρα με -στόχο. Συνεπώς 2 επιπλέον κόστος 2 (π.χ. για το A σε T)
Τότε το κόστος θα είναι 6 (2 για την 1η αλλαγή, 4 για τις άλλες)



(4) Το cepstrum ενός $v(n)$ σήματος ορίζεται ως εξής:

$$\hat{v}(n) = Z^{-1}\left[\log\left(V(z)\right)\right]$$

Έχουμε $V(z) = \dfrac{1}{\prod_{k=1}^{q}(1-c_k z^{-1})(1-c_k^{*} z^{-1})}$

Άρα
$$\log V(z) = \log\left(\frac{1}{\prod_{k=1}^{q}(1-c_k z^{-1})(1-c_k^{*} z^{-1})}\right) = -\log\left[\prod_{k=1}^{q}(1-c_k z^{-1})(1-c_k^{*} z^{-1})\right] =$$
$$= -\sum_{k=1}^{q}\log\left[(1-c_k z^{-1})(1-c_k^{*} z^{-1})\right] = -\sum_{k=1}^{q}\left[\log(1-c_k z^{-1}) + \log(1-c_k^{*} z^{-1})\right] =$$

Ξέρουμε ότι $\log(1-x) = -\sum_{n=1}^{\infty}\frac{x^n}{n}$, οπότε

$$\log V(z) = -\sum_{k=1}^{q}\left[-\sum_{n=1}^{\infty}\frac{(c_k z^{-1})^n}{n} - \sum_{n=1}^{\infty}\frac{(c_k^{*} z^{-1})^n}{n}\right] = \sum_{k=1}^{q}\sum_{n=1}^{\infty}\frac{c_k^n + c_k^{*n}}{n} z^{-n} =$$
$$= \sum_{n=1}^{\infty}\sum_{k=1}^{q}\frac{c_k^n + c_k^{*n}}{n} z^{-n} = \sum_{n=1}^{\infty}\underbrace{\sum_{k=1}^{q}\frac{r_k^n e^{j\theta_k n} + r_k^n e^{-j\theta_k n}}{n}}_{\hat{v}[n]} z^{-n}$$

Είναι σε μορφή $\sum v[n] z^{-n}$, οπότε μπορούμε να εφαρμόσουμε Z^{-1} κατευθείαν

$$\hat{v}(n) = \sum_{k=1}^{q}\frac{r_k^n e^{j\theta_k n} + r_k^n e^{-j\theta_k n}}{n} = \sum_{k=1}^{q}\frac{r_k^n}{n}\left(e^{j\theta_k n} + e^{-j\theta_k n}\right) = \sum_{k=1}^{q}\frac{r_k^n}{n}\cos(\theta_k n),\ n>0$$

(5)

Έχουμε $R_x(k) = \sum_{n=0}^{N-1-k} x(n)x(n+k)$, $R_y(k) = \sum_{n=0}^{N-1-k} y(n)y(n+k)$ συναρτήσεις αυτοσυσχέτισης

1) $$E_x = \sum_{n=0}^{N-1+p} e^2(n) = \sum_{n=0}^{N-1+p} \left(\sum_{k=0}^{p} a_{xk}\, x(n-k) \right)^2 =$$

$$= \sum_{n=0}^{N-1+p} \left(\sum_{k=0}^{p} a_{xk} x(n-k) \right) \left(\sum_{\ell=0}^{p} a_{x\ell} x(n-\ell) \right) = \sum_{k=0}^{p} a_{xk} \sum_{\ell=0}^{p} a_{x\ell} \sum_{n=0}^{N-1+p} x(n-k)x(n-\ell)$$

Ξέρουμε ότι $\sum_{n=0}^{N-1+p} x(n-k)x(n-\ell) = \sum_{n=0}^{N-1+p} x(n)x(n-\ell+k) = R_x(k-\ell)$

Συνεπώς έχουμε: $E_x = \sum_{k=0}^{p} a_{xk} \sum_{\ell=0}^{p} a_{x\ell} R_x(k-\ell)$, R_x: $p+1 \times p+1$ διαστάσεις

Άρα $E_x = a_x R_x a_x^T$

2) Ακολουθώντας την ίδια διαδικασία με πριν, καταλήγουμε στη μορφή:

$$E_{xy} = \sum_{k=0}^{p} a_{yk} \sum_{\ell=0}^{p} a_{y\ell} \sum_{n=0}^{N-1+p} x(n-k)x(n-\ell) = \sum_{k=0}^{p} \sum_{\ell=0}^{p} a_{yk} \cdot a_{y\ell} R_x(k-\ell)$$

Συνεπώς $E_{xy} = a_y R_x a_y^T$

3) Οι βέλτιστοι συντελεστές για την πρόβλεψη του $x(n)$ είναι $a_x = (a_{x0}, \ldots a_{xp})$
Αυτούς χρησιμοποιούμε στο E_x, οπότε το E_x το ελάχιστο δυνατό.
Άρα, όταν για την πρόβλεψη του $x(n)$ χρησιμοποιούμε τους a_y, τότε το σφάλμα E_{xy} θα είναι μεγαλύτερο.

Συνεπώς $E_{xy} / E_x \geq 1$

(7)

$h_t = \sigma(W h_{t-1} + U x_t)$ και $L = \sum_{t=0}^{T} L_t$

α) Έχουμε $y = \sigma(Wx)$, $y \in \mathbb{R}^d$, $x \in \mathbb{R}^d$, $W \in \mathbb{R}^{n \times d}$

$$\frac{\partial y}{\partial x} = \frac{\partial \sigma(Wx)}{\partial x} = \frac{\partial \sigma(Wx)}{\partial Wx} \cdot \frac{\partial Wx}{\partial x} = \frac{\partial \sigma(Wx)}{\partial Wx} \cdot W$$

Για τη διάφορα x_i έχουμε:

$$\frac{\partial \sigma(Wx)}{\partial Wx} = \left[\frac{\partial \sigma(Wx)}{\partial Wx_1} \cdots \frac{\partial \sigma(Wx)}{\partial Wx_k} \right] = \begin{bmatrix} \frac{\partial \sigma(Wx_1)}{\partial Wx_1} & \frac{\partial \sigma(Wx_2)}{\partial Wx_2} & \cdots & \frac{\partial \sigma(Wx_1)}{\partial Wx_n} \\ \vdots & & & \vdots \\ \frac{\partial \sigma(Wx_n)}{\partial Wx_1} & & \cdots & \frac{\partial \sigma(Wx_n)}{\partial Wx_n} \end{bmatrix}$$

Προφανώς $\frac{\partial \sigma(Wx_i)}{\partial Wx_j} = 0$ για $i \neq j$. Συνεπώς $\frac{\partial \sigma(Wx)}{\partial Wx} = \begin{bmatrix} \frac{\partial \sigma(Wx_1)}{\partial Wx_1} & 0 & \cdots & 0 \\ 0 & \frac{\partial \sigma(Wx_2)}{\partial Wx_2} & & \vdots \\ \vdots & & \ddots & \frac{\partial \sigma(Wx_n)}{\partial Wx_n} \\ 0 & \cdots & & \frac{\partial \sigma(Wx_n)}{\partial Wx_n} \end{bmatrix}$

Άρα $\frac{\partial y}{\partial x} = \frac{\partial \sigma(Wx)}{\partial Wx} \cdot W = diag(\sigma') \cdot W$

β) $\frac{\partial L}{\partial W} = \frac{\partial}{\partial W}\left(\sum_{t=0}^{T} L_t \right) = \sum_{t=0}^{T} \frac{\partial L_t}{\partial W}$ με L_t να εξαρτάται από y_t και h_t

Από σχήμα συμπεραίνουμε ότι $y_t = h_t$, συνεπώς:

$\frac{\partial L_t}{\partial W} = \frac{\partial L_t}{\partial y_t} \cdot \frac{\partial y_t}{\partial h_t} \cdot \frac{\partial h_t}{\partial W}$, με $h_t = \sigma(W h_{t-1} + U x_t)$, άρα εξάρτηση από h_{t-1}

Άρα $\frac{\partial h_t}{\partial W} = \frac{\partial h_t}{\partial h_{t-1}} \cdot \frac{\partial h_{t-1}}{\partial W}$. Παρατηρούμε ότι το h_t εξαρτάται (λόγω αναδρομής) από όλες τις χρονικές στιγμές $t-1, t-2, \dots, 1$

Συνεπώς μπορούμε να γράψουμε: $\frac{\partial L_t}{\partial W} = \sum_{k=1}^{t} \frac{\partial L_t}{\partial h_t} \cdot \frac{\partial h_t}{\partial h_k} \cdot \frac{\partial h_k}{\partial W}$ με $\frac{\partial h_t}{\partial h_k} = \prod_{j=k+1}^{t} \frac{\partial h_j}{\partial h_{j-1}}$

Άρα τελικά: $$\frac{\partial L}{\partial W} = \sum_{t=0}^{T} \sum_{k=1}^{t} \frac{\partial L_t}{\partial h_t} \frac{\partial h_t}{\partial h_k} \frac{\partial h_k}{\partial W}$$

(8)

α) Ο ρόλος της πύλης f_t είναι να καθορίσει σε τι βαθμό θα "ξεχάσουμε" - αγνοήσουμε την προηγούμενη κατάσταση του cell C_{t-1}.
Συγκεκριμένα, παράγει τιμές 0 έως 1 για κάθε στοιχείο του cell, όπου:
0: το αγνοούμε πλήρως
1: το κρατάμε πλήρως
Αυτή η πύλη είναι πολύ χρήσιμη όταν θέλουμε να ανανεώσουμε το context που έχουμε για το παρόν κομμάτι της πρότασης/κειμένου, ή για να το κρατήσουμε ίδιο για μελλοντική χρήση.

Ο ρόλος της πύλης i_t είναι να καθορίσει σε τι βαθμό οι candidate values $\tilde{C}_t$ που παράγονται, θα επηρεάσουν το τελικό C_t. Λειτουργεί σαν "scaler σημαντικότητας", για αυτό παίρνουν τιμές 0 έως 1 (από sigmoid).

Ο ρόλος της πύλης o_t είναι "φιλτράρει" το παραγόμενο C_t, ώστε στο τελικό αποτέλεσμα h_t να περάσουν συγκεκριμένα κομμάτια του C_t στην έξοδο.

β) Οι τιμές των f_t, i_t, o_t είναι πάντα στο εύρος $[0,1]$, καθώς προκύπτουν από την σιγμοειδή συνάρτηση.
Τα στοιχεία των $\tilde{C}_t$, C_t και h_t παίρνουν οποιεσδήποτε τιμές.

c) Το vanishing gradient problem προκαλείται από τον όρο $\partial C_t/\partial C_{t-1}$, ο οποίος (στα κλασσικά RNNs) είναι ανάλογος της σ' (παράγωγος της σιγμοειδής) κλπ.
και πολλαπλασιαζόμενος σε πολλά layers οδηγεί σε vanishing gradient.
Τα LSTMs καταφέρνουν να αντιμετωπίσουν το πρόβλημα αυτό, χάρη στην ιδιαίτερη αρχιτεκτονική τους.
Συνεπώς, η εξάρτηση από f_t, i_t, $\tilde{C}_t$ είναι σημαντική, καθώς η συμπεριφορά αυτών των όρων μέσα στο χρόνο λύνει ή διατηρεί το vanishing gradient problem.

δ) $\frac{\partial C_t}{\partial C_k} = \prod_{l=k+1}^{t} \frac{\partial C_l}{\partial C_{l-1}}$ με $f_t=1$ και $i_t=0$

$$C_t = f_t \odot C_{t-1} + i_t \odot \tilde{C}_t = C_{t-1}$$

$$\frac{\partial C_t}{\partial C_k} = \overbrace{\frac{\partial C_t}{\partial C_{t-1}}}^{1} \cdot \frac{\partial C_{t-1}}{\partial C_{t-2}} \cdots \frac{\partial C_{k+1}}{\partial C_k} = \frac{\partial C_{t-1}}{\partial C_{t-2}} \cdots \frac{\partial C_{k+1}}{\partial C_k}$$

Αν θεωρήσουμε ότι για κάθε t': $f_{t'}=1$ και $i_{t'}=0$, τότε $C_{t'}=C_{t'-1}$

Συνεπώς προκύπτει $\frac{\partial C_t}{\partial C_k} = 1 \cdot 1 \cdots 1 = 1$

Χωρίς αυτή την υπόθεση, πρέπει ότι για κάθε $C_{t'}$ θα πρέπει να υπολογιστεί στη γενική μορφή, όπως θα δείξουμε στο επόμενο ερώτημα.

ε) Έχουμε ότι $C_t = f_t \odot C_{t-1} + i_t \odot \tilde{C}_t \Rightarrow \frac{\partial C_t}{\partial C_{t-1}} = \frac{\partial}{\partial C_{t-1}}(f_t \odot C_{t-1}) + \frac{\partial}{\partial C_{t-1}}(i_t \odot \tilde{C}_t)$

• $\frac{\partial}{\partial C_{t-1}}(f_t \odot C_{t-1}) = \frac{\partial f_t}{\partial C_{t-1}} \odot C_{t-1} + f_t \odot \overbrace{\frac{\partial C_{t-1}}{\partial C_{t-1}}}^{1}$

Αναλύουμε τον όρο $\frac{\partial f_t}{\partial C_{t-1}} = \frac{\partial}{\partial C_{t-1}} \sigma(W_f h_{t-1} + U_f x_{t-1})$, με $h_{t-1} = o_{t-1} \odot \tanh(C_{t-1})$

Άρα $\frac{\partial f_t}{\partial C_{t-1}} = \sigma'() \cdot \frac{\partial}{\partial C_{t-1}}(W_f \cdot o_{t-1} \odot \tanh(C_{t-1})) = \sigma'() \cdot o_{t-1} \odot \frac{\partial}{\partial C_{t-1}} \tanh(C_{t-1}) \cdot W_f$

$\Rightarrow \frac{\partial f_t}{\partial C_{t-1}} = \sigma'() \cdot o_{t-1} \odot \tanh'(C_{t-1}) \cdot W_f$, αφού o_{t-1} ανεξάρτητο του C_{t-1}

• $\frac{\partial}{\partial C_{t-1}}(i_t \odot \tilde{C}_t) = \frac{\partial i_t}{\partial C_{t-1}} \odot \tilde{C}_t + i_t \odot \frac{\partial}{\partial C_{t-1}} \tilde{C}_t$, $i_t = \sigma(W_i h_{t-1} + U_i x_t)$, $\tilde{C}_t = \tanh(W_c h_{t-1} + U_c x_t)$

Ομοίως $\frac{\partial i_t}{\partial C_{t-1}} = \sigma'() \cdot \frac{\partial}{\partial C_{t-1}} W_i h_{t-1} = \sigma'() \cdot W_i \cdot o_{t-1} \odot \tanh'(C_{t-1})$

$\frac{\partial \tilde{C}_t}{\partial C_{t-1}} = \tanh'() \cdot \frac{\partial}{\partial C_{t-1}} W_c h_{t-1} = \tanh'() \cdot W_c \cdot o_{t-1} \odot \tanh'(C_{t-1})$

Ορίζοντας $\delta = o_{t-1} \odot \tanh'(C_{t-1})$ έχουμε:

$$\boxed{\frac{\partial C_t}{\partial C_{t-1}} = \sigma'() W_f \cdot \delta \odot C_{t-1} + f_t + \sigma'() W_i \delta \odot \tilde{C}_t + i_t \odot \tanh'() W_c \delta}$$

(6)

1) Έχουμε τα $\delta_t(1) = \max_{q_1 \dots q_{t-1}} P[q_1 \dots q_{t-1}, q_t = 1, O_1 \dots O_t | \lambda] =$

τμήμα A   τμήμα B

$= \max [\, P(q_1) P(q_2|q_1) \dots P(q_t = i | q_{t-1}) \cdot P(O_1|q_1) \dots P(O_t|q_t)] =$

$= \max [\, \pi_1 a_{q_2} a_{q_3} \dots P(q_t = 1 | q_{t-1}), b_{q_1}(O_1) \dots b_{q_t}(O_t)] =$

Έχουμε observations $O_1 = v, O_2 = V, O_3 = v$, $O = (v V v)$

Αρχικοποίηση: $\delta_1(1) = \pi_1 \cdot b_{q_1}(O_1)$

Άρα $\delta_1(1) = \pi_1 \cdot b_{q_1}(O_1) = 0{,}25 \cdot 0{,}5 = 0{,}125$

$\delta_1(2) = \pi_2 \cdot b_{q_2}(O_1) = 0{,}25 \cdot 0{,}2 = 0{,}05$

$\delta_1(3) = \pi_3 \cdot b_{q_3}(O_1) = 0{,}25 \cdot 0{,}75 = 0{,}1875$

$\delta_1(4) = \pi_4 \cdot b_{q_4}(O_1) = 0{,}25 \cdot 0{,}8 = \underline{0{,}2}$

Επανάληψη: $\delta_{t+1}(j) = \max_i \left(\delta_t(i) \cdot a_{ij} \cdot b_{q_j}(O_{t+1}) \right)$

• $t = 2$, $O_2 = V$:

$j=1$: $\delta_2(1) = \delta_1(1) \cdot a_{11} \cdot b_{q_1}(O_2) = 0{,}125 \cdot 0{,}25 \cdot 0{,}5 = 0{,}0156$

$\delta_2(1) = \delta_1(2) \cdot a_{21} \cdot b_{q_1}(O_2) = 0{,}05 \cdot 0{,}2 \cdot 0{,}5 = 0{,}005$ $\xrightarrow{\max}$ $\underline{\delta_2(1) = 0{,}0375}$

$\delta_2(1) = \delta_1(3) \cdot a_{31} \cdot b_{q_1}(O_2) = 0{,}1875 \cdot 0{,}4 \cdot 0{,}5 = 0{,}0375$

$\delta_2(1) = \delta_1(4) \cdot a_{41} \cdot b_{q_1}(O_2) = 0{,}2 \cdot 0{,}25 \cdot 0{,}5 = 0{,}025$

$j=2$: $\delta_2(2) = \delta_1(1) \cdot a_{12} \cdot b_{q_2}(O_2) = 0{,}125 \cdot 0{,}2 \cdot 0{,}8 = 0{,}02$

$\delta_2(2) = \delta_1(2) \cdot a_{22} \cdot b_{q_2}(O_2) = 0{,}05 \cdot 0{,}25 \cdot 0{,}8 = 0{,}01$ $\xrightarrow{\max}$ $\underline{\delta_2(2) = 0{,}0486}$

$\delta_2(2) = \delta_1(3) \cdot a_{32} \cdot b_{q_2}(O_2) = 0{,}1875 \cdot 0{,}2 \cdot 0{,}8 = 0{,}03$

$\delta_2(2) = \delta_1(4) \cdot a_{42} \cdot b_{q_2}(O_2) = 0{,}2 \cdot 0{,}3 \cdot 0{,}8 = 0{,}048$

$j=3$: $\delta_2(3) = \delta_1(1) \cdot a_{13} \cdot b_{q_3}(O_2) = 0{,}125 \cdot 0{,}3 \cdot 0{,}25 = 9{,}37 \cdot 10^{-3}$

$\delta_2(3) = \delta_1(2) \cdot a_{23} \cdot b_{q_3}(O_2) = 0{,}05 \cdot 0{,}3 \cdot 0{,}25 = 3{,}75 \cdot 10^{-3}$ $\xrightarrow{\max}$ $\underline{\delta_2(3) = 0{,}01}$

$\delta_2(3) = \delta_1(3) \cdot a_{33} \cdot b_{q_3}(O_2) = 0{,}1875 \cdot 0{,}2 \cdot 0{,}25 = 9{,}37 \cdot 10^{-3}$

$\delta_2(3) = \delta_1(4) \cdot a_{43} \cdot b_{q_3}(O_2) = 0{,}2 \cdot 0{,}2 \cdot 0{,}25 = \underline{0{,}01}$

• $j=4$: $\delta_2(4) = \delta_1(1)\cdot a_{14}\cdot b_{q_4}(O_2) = 0{,}125\cdot 0{,}25\cdot 0{,}2 = 6{,}25\cdot 10^{-3}$

$\delta_2(4) = \delta_1(2)\cdot a_{24}\cdot b_{q_4}(O_2) = 0{,}05\cdot 0{,}25\cdot 0{,}2 = 2{,}5\cdot 10^{-3}$ $\xrightarrow{max}$ $\underline{\delta_2(4) = 0{,}01}$

$\delta_2(4) = \delta_1(3)\cdot a_{34}\cdot b_{q_4}(O_2) = 0{,}1875\cdot 0{,}2\cdot 0{,}2 = 7{,}5\cdot 10^{-3}$

$\delta_2(4) = \delta_1(4)\cdot a_{44}\cdot b_{q_4}(O_2) = 0{,}2\cdot 0{,}25\cdot 0{,}2 = 0{,}01$

• $t=3$, $O_3 = U$

$j=1$: $\delta_3(1) = \delta_2(1)\, a_{11}\, b_{q_1}(O_3) = 0{,}0375\cdot 0{,}25\cdot 0{,}5 = 4{,}68\cdot 10^{-3}$

$\delta_3(1) = \delta_2(2)\cdot a_{21}\, b_{q_1}(O_3) = 0{,}048\cdot 0{,}2\cdot 0{,}5 = 4{,}8\cdot 10^{-3}$ $\xrightarrow{max}$ $\underline{\delta_3(1) = 0{,}0048}$

$\delta_3(1) = \delta_2(3)\cdot a_{31}\, b_{q_1}(O_3) = 0{,}01\cdot 0{,}4\cdot 0{,}5 = 2\cdot 10^{-3}$

$\delta_3(1) = \delta_2(4)\cdot a_{41}\, b_{q_1}(O_3) = 0{,}01\cdot 0{,}25\cdot 0{,}5 = 1{,}25\cdot 10^{-3}$

$j=2$: $\delta_3(2) = \delta_2(1)\cdot a_{12}\, b_{q_2}(O_3) = 0{,}0375\cdot 0{,}2\cdot 0{,}2 = 1{,}5\cdot 10^{-3}$

$\delta_3(2) = \delta_2(2)\cdot a_{22}\, b_{q_2}(O_3) = 0{,}048\cdot 0{,}25\cdot 0{,}2 = 2{,}4\cdot 10^{-3}$ $\xrightarrow{max}$ $\underline{\delta_3(2) = 0{,}0024}$

$\delta_3(2) = \delta_2(3)\cdot a_{32}\cdot b_{q_2}(O_3) = 0{,}01\cdot 0{,}2\cdot 0{,}2 = 4\cdot 10^{-4}$

$\delta_3(2) = \delta_2(4)\cdot a_{42}\cdot b_{q_2}(O_3) = 0{,}01\cdot 0{,}3\cdot 0{,}2 = 6\cdot 10^{-4}$

$j=3$: $\delta_3(3) = \delta_2(1)\cdot a_{13}\cdot b_{q_3}(O_3) = 0{,}0375\cdot 0{,}3\cdot 0{,}75 = 8{,}4\cdot 10^{-3}$

$\delta_3(3) = \delta_2(2)\cdot a_{23}\cdot b_{q_3}(O_3) = 0{,}048\cdot 0{,}3\cdot 0{,}75 = 0{,}0108$ $\xrightarrow{max}$ $\underline{\delta_3(3) = 0{,}0108}$

$\delta_3(3) = \delta_2(3)\cdot a_{33}\cdot b_{q_3}(O_3) = 0{,}01\cdot 0{,}2\cdot 0{,}75 = 1{,}5\cdot 10^{-3}$

$\delta_3(3) = \delta_2(4)\cdot a_{43}\cdot b_{q_3}(O_3) = 0{,}01\cdot 0{,}2\cdot 0{,}75 = 1{,}5\cdot 10^{-3}$

$j=4$: $\delta_3(4) = \delta_2(1)\cdot a_{14}\cdot b_{q_4}(O_3) = 0{,}0375\cdot 0{,}25\cdot 0{,}8 = 7{,}5\cdot 10^{-3}$

$\delta_3(4) = \delta_2(2)\cdot a_{24}\cdot b_{q_4}(O_3) = 0{,}048\cdot 0{,}25\cdot 0{,}8 = 9{,}6\cdot 10^{-3}$ $\xrightarrow{max}$ $\underline{\delta_3(4) = 0{,}0096}$

$\delta_3(4) = \delta_2(3)\cdot a_{34}\cdot b_{q_4}(O_3) = 0{,}01\cdot 0{,}2\cdot 0{,}8 = 1{,}6\cdot 10^{-3}$

$\delta_3(4) = \delta_2(4)\cdot a_{44}\cdot b_{q_4}(O_3) = 0{,}01\cdot 0{,}25\cdot 0{,}8 = 2\cdot 10^{-3}$

2) Για να καθορίσουμε την πιο πιθανή ακολουθία καταστάσεων $Q^* = (q_1, q_2, q_3)$ ακολουθούμε τον αλγόριθμο Viterbi.
Συγκεκριμένα, βρίσκουμε ποια είναι η πιθανότερη κατάσταση στο τελευταίο observation-βήμα, και έπειτα με backtracking βρίσκουμε τις πιθανότερες καταστάσεις των προηγούμενων observations.
Έτσι προκύπτει:

$$q_3 = 3, \text{ αφού } \max_i(\delta_3(i)) = 3 \quad (\delta_3(3) = 0{,}0108)$$
$$q_2 = 2, \text{ αφού } \max_i(\delta_2(i)) = 2 \quad (\delta_2(2) = 0{,}048)$$
$$q_1 = 4, \text{ αφού } \max_i(\delta_1(i)) = 4 \quad (\delta_1(4) = 0{,}2)$$

Άρα $Q^* = (3, 2, 4)$

3) $P^*(O, Q^* | \lambda) = P^*((\upsilon V \upsilon), (4,2,3))$

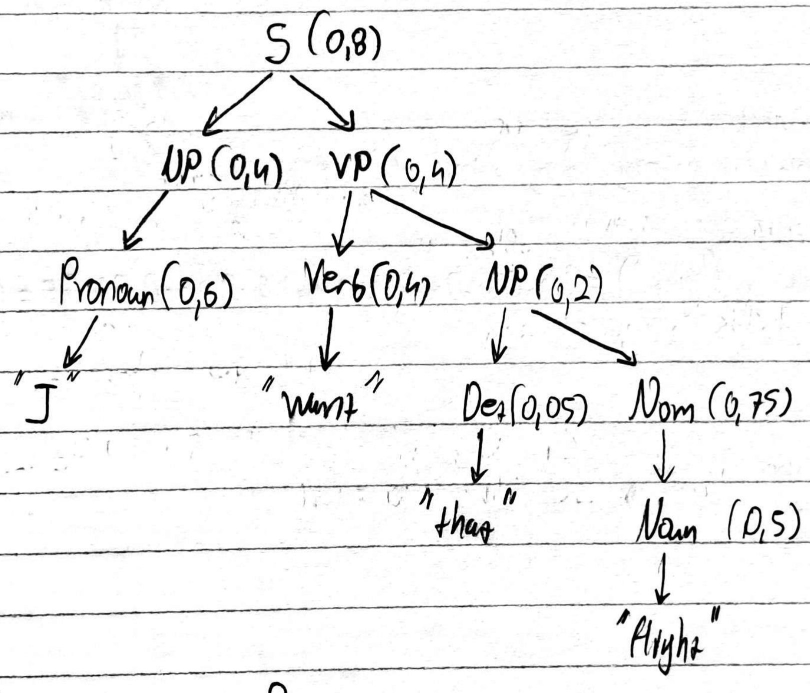
Η παραπάνω πιθανότητα είναι αυτή όπου έχουμε την ακολουθία observation O και την επιλογή καταστάσεων Q^*.
Αυτή η πιθανότητα έχει ήδη υπολογιστεί από το Viterbi και είναι η $\delta_3(3)$

Συνεπώς $P^*(O, Q^* | \lambda) = \delta_3(3) = 0{,}0108$

(9)

α) Για να έχουμε πλήρης γραμματική ως προς κάποιο κανόνα, θα πρέπει ο κανόνας αυτός να οδηγεί σε όλα τα τερματικά σύμβολα.
Συνεπώς, η γραμματική είναι πλήρης ως προς τους κανόνες που έχουν αριστερά το S, καθώς το S είναι το αρχικό σύμβολο και οδηγεί σε όλα τα τερματικά σύμβολα.
Αντίθετα, η γραμματική δεν είναι πλήρης ως προς τους κανόνες που έχουν αριστερά το NP, καθώς από το NP δεν πάμε σε όλα τα τερματικά σύμβολα.
Συγκεκριμένα από το NP δεν μπορούμε να οδηγηθούμε στο Verb, με συνέπεια να μην φτάνουμε ποτέ στο τερματικό want.
Επίσης, αθροίζοντας τις πιθανότητες των κανόνων με αριστερό σύμβολο το NP έχουμε άθροισμα 1, που σημαίνει ότι δεν υπάρχουν άλλοι κανόνες με NP στο αριστερό μέρος. Αντίθετα, οι πιθανότητες των κανόνων με αριστερό σύμβολο S δεν αθροίζουν στο 1

β) Παρακάτω το δέντρο που παράγει την πρόταση "I want that flight":

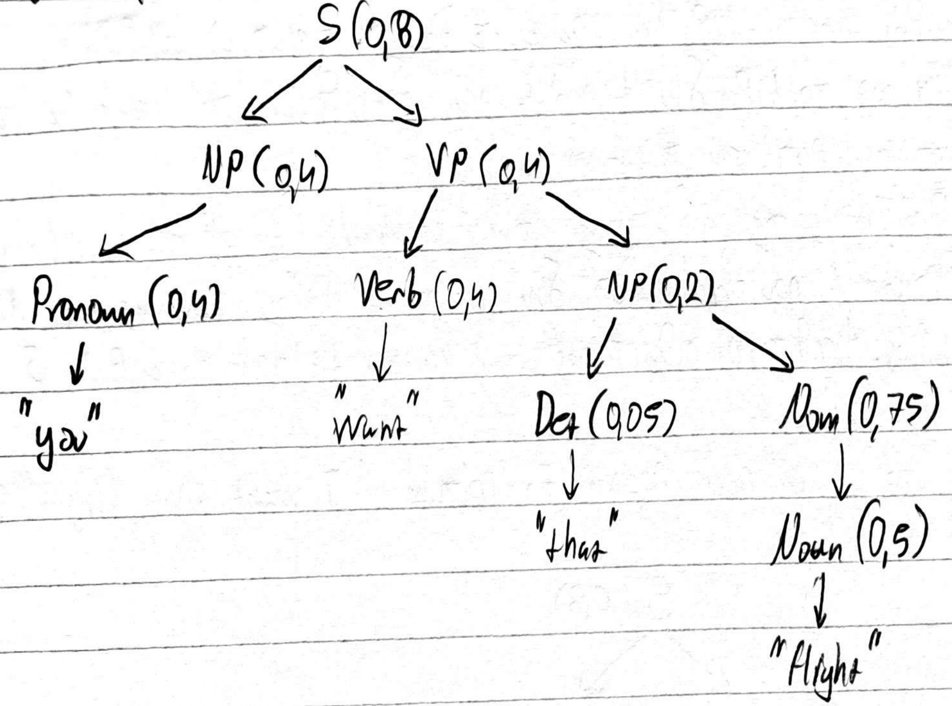


Δεν έχουμε αμφισημία (ambiguity) καθώς:
1) Για να έχω "I" πρέπει να έχω Pronoun, στο οποίο οδηγούμε μόνο από NP
2) Για να έχω "want" πρέπει να έχω Verb, στο οποίο οδηγούμε από VP με 2 τρόπους.
Δεδομένου όμως ότι η πρόταση δεν τελειώνει στο "want", ο άλλος κανόνας απορρίπτεται
3) Τα "that" και "flight" ομοίως παράγονται μοναδικά από Det και Nom-Noun

γ) Για να υπολογιστεί η πιθανότητα του δέντρου της "I want that flight" έχουμε:

$$P(\text{"I want that flight"}) = 0,8 \cdot 0,4 \cdot 0,4 \cdot 0,6 \cdot 0,4 \cdot 0,2 \cdot 0,05 \cdot 0,75 \cdot 0,5 = 0,0001152$$

Για την πιθανότητα της πρότασης "you want that flight" θα σχεδιάσουμε πρώτα το συντακτικό δέντρο:



Οπότε εδώ έχουμε αφαιρέσει ότι τους ίδιους όρους με πριν.
Η πιθανότητα εδώ είναι:

$$P(\text{"you want a flight"}) = 0,8 \cdot 0,4 \cdot 0,4 \cdot 0,4 \cdot 0,4 \cdot 0,2 \cdot 0,05 \cdot 0,75 \cdot 0,5 = 0,0000768$$