Programmation dynamique pour HMM

- Le calcul des $\alpha(i,t)$ donne un balayage de gauche à droite
- On peut faire la même chose, mais de droite à gauche
  - on définit $\beta(i,t) = P(S_{t+1:T}=s_{t+1:T} \mid H_t = i)$
  - on note la récursion

    $$\begin{aligned}\beta(i,t\text{-}1) &= P(S_{t:T}=s_{t:T} \mid H_{t-1} = i)\\ &= \textstyle\sum_j P(S_{t:T}=s_{t:T}, H_t = j \mid H_{t-1} = i) = \sum_j P(S_{t+1:T}=s_{t+1:T}, S_t= s_t, H_t = j \mid H_{t-1} = i)\\ &= \textstyle\sum_j P(S_t= s_t \mid H_t = j)\, P(H_t = j \mid H_{t-1} = i)\, P(S_{t+1:T}=s_{t+1:T} \mid H_t = j)\\ &= \textstyle\sum_j P(S_t= s_t \mid H_t = j)\, P(H_t = j \mid H_{t-1} = i)\, \beta(j,t)\end{aligned}$$
  - on a les valeurs initiales $\beta(i,T) = 1\ \forall i$

- Une fois le tableau β calculé, on obtient facilement:

$$\begin{aligned}P(S_{1:T}=s_{1:T}) &= \textstyle\sum_j P(S_{1:T}=s_{1:T}, H_1 = j)\\ &= \textstyle\sum_j P(S_{2:T}=s_{2:T} \mid H_1 = j)\, P(S_1=s_1 \mid H_1 = j)\, P(H_1 = j)\\ &= \textstyle\sum_j \beta(j,1)\, P(S_1=s_1 \mid H_1 = j)\, P(H_1 = j)\end{aligned}$$

Programmation dynamique pour HMM

- Exemple: décoder un message binaire avec canal bruité (T=4)
  - message observé: S_1=0, S_2=0, S_3=0, S_4=1

Modèle d'observation

	H_t=0	H_t=1
$P(S_t=0 \mid H_t)$	0.9	0.2
$P(S_t=1 \mid H_t)$	0.1	0.8

Modèle de transition

	H_{t-1}=0	H_{t-1}=1
$P(H_t=0 \mid H_{t-1})$	0.3	0.6
$P(H_t=1 \mid H_{t-1})$	0.7	0.4

Distribution initiale

	H_1=0	H_1=1
$P(H_1)$	0.5	0.5

β(i,t)

i \ t	1	2	3	4
0				
1				

  - initialisation: β(i,4) = 1

Hugo Larochelle et Froduald Kabanza

Programmation dynamique pour HMM

- Exemple: décoder un message binaire avec canal bruité (T=4)
  - message observé: S_1=0, S_2=0, S_3=0, S_4=1

Modèle d'observation

	H_t=0	H_t=1
$P(S_t=0 \mid H_t)$	0.9	0.2
$P(S_t=1 \mid H_t)$	0.1	0.8

Modèle de transition

	H_{t-1}=0	H_{t-1}=1
$P(H_t=0 \mid H_{t-1})$	0.3	0.6
$P(H_t=1 \mid H_{t-1})$	0.7	0.4

Distribution initiale

	H_1=0	H_1=1
$P(H_1)$	0.5	0.5

β(i,t)

i \ t	1	2	3	4
0				1
1				1

- initialisation: β(i,4) = 1

Programmation dynamique pour HMM

- Exemple: décoder un message binaire avec canal bruité (T=4)
  - message observé: S_1=0, S_2=0, S_3=0, S_4=1

Modèle d'observation

	H_t=0	H_t=1
$P(S_t=0 \mid H_t)$	0.9	0.2
$P(S_t=1 \mid H_t)$	0.1	0.8

Modèle de transition

	H_{t-1}=0	H_{t-1}=1
$P(H_t=0 \mid H_{t-1})$	0.3	0.6
$P(H_t=1 \mid H_{t-1})$	0.7	0.4

Distribution initiale

	H_1=0	H_1=1
$P(H_1)$	0.5	0.5

β(i,t)

i \ t	1	2	3	4
0				1
1				1

  - récursion (t=4): $\beta(i,t-1) = \sum_j P(S_t = s_t \mid H_t = j)\, P(H_t = j \mid H_{t-1} = i)\, \beta(j,t)$

Programmation dynamique pour HMM

- Exemple: décoder un message binaire avec canal bruité (T=4)
  - message observé: S_1=0, S_2=0, S_3=0, S_4=1

Modèle d'observation

	H_t=0	H_t=1
$P(S_t=0 \mid H_t)$	0.9	0.2
$P(S_t=1 \mid H_t)$	0.1	0.8

Modèle de transition

	H_{t-1}=0	H_{t-1}=1
$P(H_t=0 \mid H_{t-1})$	0.3	0.6
$P(H_t=1 \mid H_{t-1})$	0.7	0.4

Distribution initiale

	H_1=0	H_1=1
$P(H_1)$	0.5	0.5

β(i,t)

i \ t	1	2	3	4
0				1
1				1

- récursion $\beta(0,3) = P(S_4=1|H_4=0)\ P(H_4=0|H_3=0)\ \beta(0,4) + P(S_4=1|H_4=1)\ P(H_4=1|H_3=0)\ \beta(1,4)$

Hugo Larochelle et Froduald Kabanza

Programmation dynamique pour HMM

- Exemple: décoder un message binaire avec canal bruité (T=4)
  - message observé: $S_1=0$, $S_2=0$, $S_3=0$, $S_4=1$

Modèle d'observation

	$H_t=0$	$H_t=1$
$P(S_t=0 \mid H_t)$	0.9	0.2
$P(S_t=1 \mid H_t)$	0.1	0.8

Modèle de transition

	$H_{t-1}=0$	$H_{t-1}=1$
$P(H_t=0 \mid H_{t-1})$	0.3	0.6
$P(H_t=1 \mid H_{t-1})$	0.7	0.4

Distribution initiale

	$H_1=0$	$H_1=1$
$P(H_1)$	0.5	0.5

β(i,t)

i \ t	1	2	3	4
0			0.59	1
1				1

  - récursion β(0,3) = 0.1 x 0.3 x 1 + 0.8 x 0.7 x 1 = 0.59

Programmation dynamique pour HMM

- Exemple: décoder un message binaire avec canal bruité (T=4)
  - message observé: S_1=0, S_2=0, S_3=0, S_4=1

Modèle d'observation

	H_t=0	H_t=1
$P(S_t=0 \mid H_t)$	0.9	0.2
$P(S_t=1 \mid H_t)$	0.1	0.8

Modèle de transition

	H_{t-1}=0	H_{t-1}=1
$P(H_t=0 \mid H_{t-1})$	0.3	0.6
$P(H_t=1 \mid H_{t-1})$	0.7	0.4

Distribution initiale

	H_1=0	H_1=1
$P(H_1)$	0.5	0.5

β(i,t)

i \ t	1	2	3	4
0			0.59	1
1				1

  - récursion β(1,3) = $P(S_4=1|H_4=0)\ P(H_4=0|H_3=1)\ \beta(0,4)$
    $+\ P(S_4=1|H_4=1)\ P(H_4=1|H_3=1)\ \beta(1,4)$

Programmation dynamique pour HMM

- Exemple: décoder un message binaire avec canal bruité (T=4)
  - message observé: S_1=0, S_2=0, S_3=0, S_4=1

Modèle d'observation

	H_t=0	H_t=1
$P(S_t=0 \mid H_t)$	0.9	0.2
$P(S_t=1 \mid H_t)$	0.1	0.8

Modèle de transition

	H_{t-1}=0	H_{t-1}=1
$P(H_t=0 \mid H_{t-1})$	0.3	0.6
$P(H_t=1 \mid H_{t-1})$	0.7	0.4

Distribution initiale

	H_1=0	H_1=1
$P(H_1)$	0.5	0.5

β(i,t)

i \ t	1	2	3	4
0			0.59	1
1			0.38	1

  - récursion β(1,3) = 0.1 x 0.6 x 1 + 0.8 x 0.4 x 1 = 0.38

Programmation dynamique pour HMM

- Exemple: décoder un message binaire avec canal bruité (T=4)
  - message observé: $S_1=0$, $S_2=0$, $S_3=0$, $S_4=1$

Modèle d'observation

	$H_t=0$	$H_t=1$
$P(S_t=0 \mid H_t)$	0.9	0.2
$P(S_t=1 \mid H_t)$	0.1	0.8

Modèle de transition

	$H_{t-1}=0$	$H_{t-1}=1$
$P(H_t=0 \mid H_{t-1})$	0.3	0.6
$P(H_t=1 \mid H_{t-1})$	0.7	0.4

Distribution initiale

	$H_1=0$	$H_1=1$
$P(H_1)$	0.5	0.5

β(i,t)

i \ t	1	2	3	4
0			0.59	1
1			0.38	1

- récursion (t=3): $\beta(i,t-1) = \sum_j P(S_t = s_t \mid H_t = j)\, P(H_t = j \mid H_{t-1} = i)\, \beta(j,t)$

Programmation dynamique pour HMM

- Exemple: décoder un message binaire avec canal bruité (T=4)
  - message observé: S_1=0, S_2=0, S_3=0, S_4=1

Modèle d'observation

	H_t=0	H_t=1
$P(S_t=0 \mid H_t)$	0.9	0.2
$P(S_t=1 \mid H_t)$	0.1	0.8

Modèle de transition

	H_{t-1}=0	H_{t-1}=1
$P(H_t=0 \mid H_{t-1})$	0.3	0.6
$P(H_t=1 \mid H_{t-1})$	0.7	0.4

Distribution initiale

	H_1=0	H_1=1
$P(H_1)$	0.5	0.5

β(i,t)

i \ t	1	2	3	4
0			0.59	1
1			0.38	1

- récursion β(0,2) = $P(S_3=0|H_3=0)\, P(H_3=0|H_2=0)\, \beta(0,3)$
  $+ P(S_3=0|H_3=1)\, P(H_3=1|H_2=0)\, \beta(1,3)$

Programmation dynamique pour HMM

- Exemple: décoder un message binaire avec canal bruité (T=4)
  - message observé: S_1=0, S_2=0, S_3=0, S_4=1

Modèle d'observation

	H_t=0	H_t=1
$P(S_t=0 \mid H_t)$	0.9	0.2
$P(S_t=1 \mid H_t)$	0.1	0.8

Modèle de transition

	H_{t-1}=0	H_{t-1}=1
$P(H_t=0 \mid H_{t-1})$	0.3	0.6
$P(H_t=1 \mid H_{t-1})$	0.7	0.4

Distribution initiale

	H_1=0	H_1=1
$P(H_1)$	0.5	0.5

β(i,t)

i \ t	1	2	3	4
0		0.2125	0.59	1
1			0.38	1

  - récursion β(0,2) = 0.9 x 0.3 x 0.59 + 0.2 x 0.7 x 0.38 = 0.2125

Programmation dynamique pour HMM

- Exemple: décoder un message binaire avec canal bruité (T=4)
  - message observé: S_1=0, S_2=0, S_3=0, S_4=1

Modèle d'observation

	H_t=0	H_t=1
$P(S_t=0 \mid H_t)$	0.9	0.2
$P(S_t=1 \mid H_t)$	0.1	0.8

Modèle de transition

	H_{t-1}=0	H_{t-1}=1
$P(H_t=0 \mid H_{t-1})$	0.3	0.6
$P(H_t=1 \mid H_{t-1})$	0.7	0.4

Distribution initiale

	H_1=0	H_1=1
$P(H_1)$	0.5	0.5

β(i,t)

i \ t	1	2	3	4
0	0.106235	0.2125	0.59	1
1	0.14267	0.349	0.38	1

  - on continue d'appliquer la récursion jusqu'au début (t=1)...

Hugo Larochelle et Froduald Kabanza

Lissage avec un HMM

- Les tables α(i,t) et β(i,t) peuvent également être utilisées pour faire du lissage
  - $P(H_k = i \mid S_{1:T}{=}s_{1:T}) = P(H_k = i,\, S_{1:k}{=}s_{1:k},\, S_{k+1:T}{=}s_{k+1:T}) \,/\, \Upsilon$ (ϒ est la normalisation)

    $= P(H_k = i,\, S_{1:k}{=}s_{1:k})\; P(S_{k+1:T}{=}s_{k+1:T} \mid H_k = i) \,/\, \Upsilon$

    $= \alpha(i,k)\; \beta(i,k) \,/\, \Upsilon$

- ϒ correspond à une somme sur *i* seulement

Lissage avec un HMM

- Exemple: décoder un message binaire avec canal bruité (T=4)
  - message observé: S_1=0, S_2=0, S_3=0, S_4=1

α(i,t)

i \ t	...	2	...
0	...	0.1755	...
1	...	0.071	...

β(i,t)

i \ t	...	2	...
0	...	0.2125	...
1	...	0.349	...

  - on peut calculer les probabilités de lissage au temps t=2

$$P(H_2 = 0 \mid S_1{=}0, S_2{=}0, S_3{=}0, S_4{=}1) = \frac{\alpha(0,2)\ \beta(0,2)}{\sum_i \alpha(i,2)\ \beta(i,2)}$$

$$= \alpha(0,2)\ \beta(0,2) / \left(\alpha(0,2)\ \beta(0,2) + \alpha(1,2)\ \beta(1,2)\right)$$

$$= 0.1755 \times 0.2125 / (0.1755 \times 0.2125 + 0.071 \times 0.349)$$

$$\approx 0.6008$$

$$P(H_2 = 1 \mid S_1{=}0, S_2{=}0, S_3{=}0, S_4{=}1) = 0.071 \times 0.349 / (0.1755 \times 0.2125 + 0.071 \times 0.349)$$

$$\approx 0.3992$$

Lissage avec un HMM

- On peut également faire du lissage sur deux variables cachées adjacentes
  - $P(H_k = i, H_{k+1} = j \mid S_{1:T}=s_{1:T})$
    $= P(H_k = i, H_{k+1} = j, S_{1:k}=s_{1:k}, S_{k+1:T}=s_{k+1:T}) \,/\, \Upsilon'$
    $= P(H_k = i, S_{1:k}=s_{1:k})\, P(H_{k+1} = j \mid H_k = i)\, P(S_{k+1}=s_{k+1} \mid H_{k+1} = j)\, P(S_{k+2:T}=s_{k+2:T} \mid H_{k+1} = j) \,/\, \Upsilon'$
    $= \alpha(i,k)\, P(H_{k+1} = j \mid H_k = i)\, P(S_{k+1}=s_{k+1} \mid H_{k+1} = j)\, \beta(j,k+1) \,/\, \Upsilon'$

- Υ' est une somme sur *i* **et** *j*