A Tutorial on Hidden Markov Models and Selected Applications in Speech Recognition

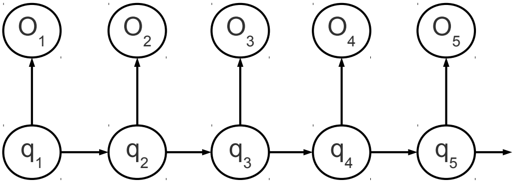
L.R. Rabiner, 1989, IEEE

Plan

1. Caractérisation d'un HMM
2. Les trois problèmes de base des HMM
3. Discussion / Limites

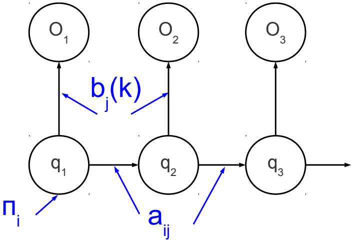
Modèles de Markov Cachés

- Données: Une série d'observations $\mathbf{O} = O_1 O_2 O_2 ... O_T$
- Modèle: Les données sont générées par un processus (une chaine) de Markov sur des états $\mathbf{Q} = q_1 q_2 q_3 ... q_T$ (cachés car non observés).



B. Eléments d'un HMM (p. 260)

- Les états cachés:
  - nombre d'états N
  - $\mathbf{S} = \{S_1, S_2, \ldots, S_N\}$
  - L'état au temps t: q_t
- Les valeurs possibles pour O_t = un alphabet discret
  - nombre de symboles: M
  - $\mathbf{V} = \{v_1, v_2, \ldots, v_M\}$
- Les probabilités de transition entre états $A = \{a_{ij}\}$:
  $a_{ij} = P(q_{t+1} = S_j | q_t = S_i)$
- Les probabilités d'émission $B = \{b_j(k)\}$:
  $b_j(k) = P(O_t = v_k | q_t = Sj)$
- Les probabilités initiales:
  $\pi_i = P(q_1 = S_i)$



Présentation des trois problèmes

Problème 1: Calcul de la vraisemblance Etant donné une séquence d'observations $\mathbf{O}$ et un HMM $\lambda = (A, B, \pi)$, comment calculer efficacement $P(\mathbf{O}|\lambda)$?

Problème 2: Détermination de la séquence optimale Etant donné une séquence d'observations $\mathbf{O}$ et le modèle λ, comment choisir la séquence $\mathbf{Q}$ qui soit optimale (*e.g.* qui explique le mieux les observations) ?

Problème 3: Estimation Comment ajuster les paramètres λ pour maximiser $P(\mathbf{O}|\lambda)$?

Calcul de Vraisemblance: Solution naive

- Sachant une séquence d'état cachés $\mathbf{Q}$:

$$P(\mathbf{O}|\mathbf{Q},\lambda) = \prod_{t=1}^{T} P(O_t|q_t,\lambda) = b_{q_1}(O1).b_{q_2}(O2)\ldots b_{q_t}(O_t)$$

- La probabilité d'une séquence d'états cachés $\mathbf{Q}$ est simplement:

$$P(\mathbf{Q}|\lambda) = \pi_{q_1} a_{q_1 q_2} a_{q_2 q_3} \ldots a_{q_{T-1} q_T}$$

- Finalement:

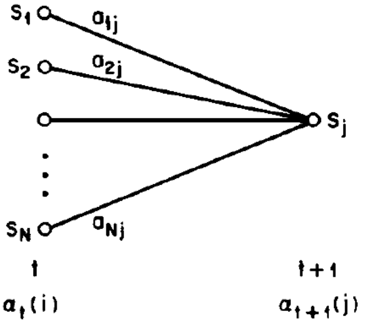
$$P(\mathbf{O}|\lambda) = \sum_{\text{all Q}} P(\mathbf{O}|\mathbf{Q},\lambda) P(\mathbf{Q}|\lambda)$$

- Mais complexité en $T.N^T$: inefficace

Calcul de Vraisemblance: Procédure Forward (p. 262)

Calcul des variables forward :

$$\alpha_t(i) = P(O_1, O_2, ...O_t, q_t = S_i|\lambda) \text{ pour } t \in [1..T]$$



1. Initialisation:

$$\alpha_1(i) = \pi_i b_i(O_1)$$

2. Induction:

$$\alpha_{t+1}(j) = \Big[\sum_{i=1}^{N} \alpha_t(i) a_{ij}\Big] b_j(O_{t+1})$$

3. Termination:

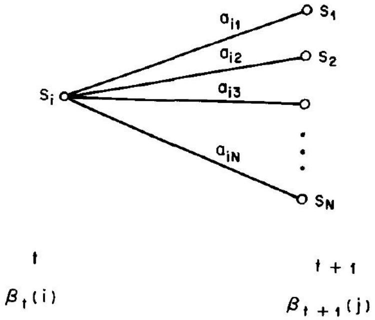
$$P(\mathbf{O}|\lambda) = \sum_{i=1}^{N} \alpha_T(i)$$

Complexité: $N^2.T$ au maximum

Procédure Backward (p. 263)

Calcul des variables forward :

$$\beta_t(i) = P(O_{t+1}, O_{t+2}, ...O_T | q_t = S_i, \lambda) \text{ pour } t \in [1..T]$$



1. Initialisation:

$$\beta_T(i) = 1$$

2. Induction:

$$\beta_t(i) = \sum_{i=1}^{N} a_{ij} b_j(O_{t+1}) \beta_{t+1}(j)$$

Complexité: $N^2.T$ au maximum

Calcul de l'état le plus probable à chaque position

- A chaque position, on peut chercher l'état S_i qui maximise:

$$\gamma_t(i) = P(q_t = S_i|\mathbf{O}, \lambda)$$

$$\gamma_t(i) = \frac{\alpha_t(i)\beta_t(i)}{\sum_{i=1}^{N} \alpha_t(i)\beta_t(i)}$$

- Maximise le nombre d'états chachés "corrects" moyens.
- Problème: la séquence obtenue peut ne pas être valide

Recherche de la séquence de probabilité maximale: Algorithme de Viterbi (p. 264)

On définit le score suivant:

$$\delta_t(i) = \max_{q_1, q_2, \ldots, q_{t-1}} P(q_1 q_2 \ldots q_t = S_i, O_1, O_2 \ldots O_t | \lambda)$$

$\delta_t(i)$ est le score max (probabilité) pour une séquence qui prend en compte les observations jusqu'à t et qui arrive à l'état S_i en t.

1. Initialisation:

$$\delta_1(i) = \pi_i b_i(O_1) \qquad \psi_1(i) = 0$$

2. Recurrence:

$$\delta_t(j) = \max_{1 \leq i \leq N} [\delta_{t-1}(i) a_{ij}] b_j(O_t) \qquad \psi_t(j) = argmax[\delta_{t-1}(i) a_{ij}]$$

3. Fin:

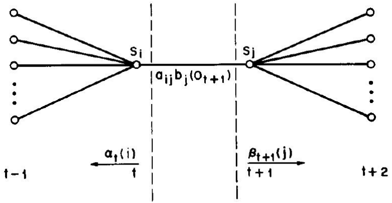
$$P^* = max[\delta_T(i)] \qquad q_T^* = argmax[\delta_T(i)]$$

4. Backtracking: $q_t^* = \psi_{t+1}(q_{t+1}^*)$

Problème

- *By far the most difficult problem of HMMs is to determine a method to adjust the model parameters (A,B,π) to maximize the [likelihood]*
- Recherche d'un optimum local en utilisant l'algorithme de Baum-Welch (algorithme EM).
- Se calcule en utilisant les variables α,β,γ précédentes, plus :

$$\xi_t(i,j) = P(q_t = S_i, q_{t+1} = S_j | \mathbf{O}, \lambda)$$



$$\xi_t(i,j) = \frac{\alpha_t(i) a_{ij} b_j(O_{t+1}) \beta_{t+1}(j)}{P(\mathbf{O}|\lambda)}$$

- Les détails d'implémentation dépendent du modèle (transitions constantes ou locales, comptage d'évènements ...) En résumé:

$$\bar{\pi}_i = \text{ expected number of times in state } S_i \text{ at time (t =1)}$$

$$\bar{a_{ij}} = \frac{\text{expected number of transitions from state } S_i \text{ to state } S_j}{\text{expected number of transitions from state } S_i}$$

$$b_j\bar{(k)} = \frac{\text{expected number of times in state j and observing symbol } v_k}{\text{expected number of times in state j}}$$

qui se calculent conditionellement aux valeurs courantes et permettent de les mettre à jour

Conclusions

- Avec l'ensemble de ces résultats, il est possible d'implémenter l'estimation, l'utilisation d'un modèle HMM simple.
- Extensions sont nombreuses (cf. exposés suivants)
- Quelques HMM célèbres en génétique:
  - **Lander et Green (1987)**: Calculs de vraisemblance dans des pedigrees complexes. Etats Cachés: indicateurs de ségrégations. Observations: génotypes. Séquence: marqueurs le long des chromosomes. (CRIMAP)
  - **Li and Stephens (2003)** : Etats cachés: haplotypes connus. Etat observé: haplotype / génotype. (PHASE)
  - **Scheet et Stephens (2006)** (fastPHASE) ; **Browning's (2005 ...)** (Beagle): Etats cachés: groupes d'haplotypes locaux. Etats observés: haplotypes / génotypes.