

Exercice 3 — solution esquissée

Soit \mathcal{M} une HMM avec

- ensemble d'états \mathcal{Q} de taille N
- probabilités de transition $\tau: \mathcal{Q} \times \mathcal{Q} \mapsto [0, 1]$ où $\sum_{q' \in \mathcal{Q}} \tau(q, q') = 1$ pour tout $q \in \mathcal{Q}$
- probabilités d'émission $p: \mathcal{Q} \times \Sigma \mapsto [0, 1]$ où $\sum_{c \in \Sigma} p(q, c) = 1$ pour tout $q \in \mathcal{Q}$.
- probabilités d'état initial $\pi: \mathcal{Q} \mapsto [0, 1]$ où $\sum_{q \in \mathcal{Q}} \pi(q) = 1$.

On a une séquence $s = s_1 \cdots s_\ell$. Pour trouver le deuxième meilleur chemin, on utilise PD comme pour le chemin de Viterbi. Soit

$$\delta_i(q) = \max_{q_1, \dots, q_{i-1}} \mathbb{P}\{q_1 \cdots q_{i-1} q_i = q, s_1 \cdots s_i\}, \quad (1)$$

la probabilité que le meilleur chemin pour le préfixe finit en état q . On définit aussi

$$\phi_i(q) = \text{second}_{q_1, \dots, q_{i-1}} \mathbb{P}\{q_1 \cdots q_{i-1} q_i = q, s_1 \cdots s_i\}, \quad (2)$$

où $\text{second } \mathcal{S}$ dénote le deuxième plus grand élément d'un ensemble fini \mathcal{S} . $\phi_i(q)$ est la probabilité que le deuxième meilleur chemin pour le préfixe de longueur i finit en état q .

Finalement, on va utiliser aussi $u_i(q) \in \mathcal{Q}$ et $v_i(q) \in \mathcal{Q}$ pour stocker les avant-derniers états en (1) et (2) pour lesquelles les maximums sont atteints.

Initialisation : Pour chaque q , $\delta_1(q) = \pi(q)p(q, s_1)$; $u_1(q) = v_1(q) = \text{nil}$; $\phi_1(q) = 0$ [pas de deuxième choix].

Récurrences : Pour chaque q ,

$$\delta_{i+1}(q) = \left(\max_{q'} \delta_i(q') \tau(q', q) \right) p(q, s_{i+1}),$$

$u_{i+1}(q)$ affecté à q' pour lequel le maximum est atteint.

$$\phi_{i+1}(q) = p(q, s_{i+1}) \max \left\{ \begin{array}{l} \max_{q'' \in \mathcal{Q} - \{u_{i+1}(q)\}} \delta_i(q'') \tau(q'', q), \\ \phi_i(u_{i+1}(q)) \tau(u_{i+1}(q), q) \end{array} \right\}.$$

Les deux termes correspondent aux possibilités (1) le chemin vient de $q'' \neq u_{i+1}(q)$ donc il est optimal pour arriver en q'' , ou (2) le chemin vient de $u_{i+1}(q)$ donc il est le 2ème meilleur pour arriver en $u_{i+1}(q)$. La valeur $v_{i+1}(q)$ est affectée à q'' où $u_{i+1}(q)$ selon le cas.

Le 2ème meilleur chemin est retrouvé à partir de l'état q_ℓ avec le maximum ϕ_ℓ , en reculant par les v_i : $q_{\ell-1} = v_\ell(q_\ell)$, $q_{\ell-2} = v_{\ell-1}(q_{\ell-1})$, ...

Temps de calcul : $O(N^2\ell)$ en général, $O(N\ell)$ pour des HMM de profile (car 3 éléments dans chaque $\max\{\dots\}$).