

4.3 Algorithme de Viterbi

Étant donné un HMM M et une séquence w générée selon M , le chemin de Viterbi de w dans M est la séquence V_w d'états cachés la plus probable permettant de la générer. Plus formellement, V_w est la séquence $c \in S^{|w|}$ qui maximise $P(c|w)$:

$$V_w = \text{ArgMax}_{c \in S^{|w|}} P(c|w).$$

D'après la formule de Bayes, on a $P(c|w) = \frac{P(w|c)P(c)}{P(w)}$. On peut donc aussi écrire

$$V_w = \text{ArgMax}_{c \in S^{|w|}} P(w|c)P(c).$$

Si $w = o_1 \dots o_n$ et $c = s_1 \dots s_n$,

$$P(w|c)P(c) = \prod_{i=1}^n E(s_i, o_i) \cdot \pi(s_1) \prod_{i=1}^{n-1} T(s_i, s_{i+1}).$$

Comme il est pratiquement impossible d'énumérer tous les chemins de longueur n , on a recours à une méthode de programmation dynamique pour rendre le problème praticable. Pour cela, on considère la séquence d'états $c_{k,i}$ la plus probable permettant de générer $o_1 \dots o_i$ et terminant dans l'état k . On note $p_{k,i}$ la probabilité associée.

Exercice 14

1. Comment calculer $c_{k,1}$? $p_{k,1}$?
2. Comment calculer $c_{k,i+1}$ et $p_{k,i+1}$ en fonction des valeurs précédentes ?
3. Écrivez la méthode `Viterbi(self, w)` qui calcule le chemin de Viterbi de w et la probabilité associée. Pour accroître la robustesse des calculs, vous pouvez calculer plutôt le logarithme de cette probabilité.

Exercice 15 Soit M un HMM, $w = o_1 \dots o_n$ une séquence générée à partir de M , $v_1 \dots v_n$ le chemin de Viterbi de w dans M et $pv(w)$ la probabilité associée. Quelle est la probabilité que w soit généré le long de son chemin de Viterbi ? Écrivez un programme permettant d'évaluer ce nombre et testez le sur les HMMs définis précédemment pour des longueurs variables de w .

Exercice 16 Soit M un HMM, $w = o_1 \dots o_n$ une séquence générée à partir de M , $s_1 \dots s_n$ la suite d'états correspondant. Soit $v_1 \dots v_n$ le chemin de Viterbi de w dans M . On s'intéresse au nombre d'indices i tels que $s_i = v_i$. Réalisez une expérience avec l'un des HMMs définis précédemment, et $n = 1000$ permettant d'évaluer ce nombre.