Modèles de Markov Cachés

Par Djellali Hayet

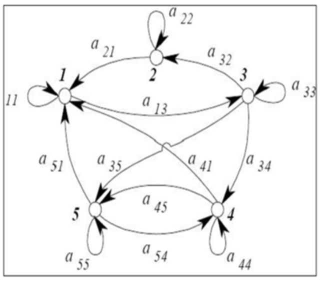
2020

HISTORIQUE

Les chaînes de Markov ont été inventées par Andreï Markov: est un mathématicien russe, Ses travaux sur la théorie des probabilités l'ont amené à mettre au point les chaînes de Markov. Il a publié les premiers résultats sur les chaînes de Markov à espace d'états fini en 1906.

La théorie des modèles de Markov cachés a été développée dans les années 1960 et début 1970 «modèle de Markov caché» a été inventé par Neuwirth.

un processus de Markov



Approche statistique en RAP

- Etant donné une trame sonore représentatives de quelques mots différents.
- Les Modèles de Markov Cachés
- Décodage optimal ?
- Le décodage optimal détermine la
- séquence W qui fournit la meilleur
- interprétation de l'observation X fournit par l'analyse acoustique.

Type de HMM

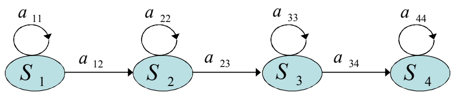
- Le Modèle ERGODIC
- Dans le Modèle ergodic il y a transition de chaque etat vers tous les autres états.
- Modele Gauche Droite (BAKIS):
- Le nombre de transition est limitée
- Il y a transition de la gauche vers la droite.

Gramaire.txt

- $word1 = sifr1 | wahid1 | ithnane1;
- $word2 = sifr2 | wahid2 | ithnane2;
- ({START_SIL } [$word1] | [$word2] {END_SIL})

HMM Gauche Droite

•La transition se fait vers l'etat suivant ou reste dans dans le même état



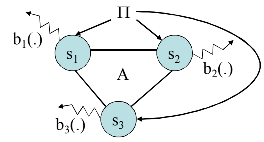
4-état d'un HMM gauche droite sans transitions avec saut

$$\mathbf{A} = \begin{bmatrix} a_{11} & a_{12} & 0 & 0 \\ 0 & a_{22} & a_{23} & 0 \\ 0 & 0 & a_{33} & a_{34} \\ 0 & 0 & 0 & a_{44} \end{bmatrix}$$

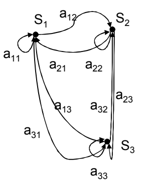
Composantes d'un MMC (« HMM »)

- **Les probabilités initiales des états** $\Pi=\{\pi_i=P(s_i)\}$
- **Le modèle de transition des états**
  - L'alphabet $\Sigma=\{s_1,\ldots,s_m\}$ décrivant les états de la chaîne de Markov
  - La matrice des probabilités de transitions entre états $A=\{a_{ij}=P(s_j|s_i)\}$
- **Le modèle d'observation de l'évidence**
  - L'alphabet $\Omega=\{o_1,\ldots,o_k\}$ des symboles émis par les s_i pour un HMM discret
  - Les probabilités d'émission $B=\{b_i(o_k)=P(o_k|s_i)\}$

On suppose un processus stationnaire (les modèles de transition et d'observation sont constants dans le temps)



Processus de Markov Discret



States=$\{S_1,S_2,S_3\}$ $P(q_t=S_i|q_{t-1}=S_j)=a_{ij}$

$Sum_{(j=1..N)}(aij)=1$

Consider the following model of weather:
- •S1: rain or snow
- •S2: cloudy
- •S3: sunny

$$A=\{a_{ij}\}=\begin{bmatrix} 0.4 & 0.3 & 0.3 \\ 0.2 & 0.6 & 0.2 \\ 0.1 & 0.1 & 0.8 \end{bmatrix}$$

What is the probability that the weather for the next 3 days will be "sun,sun,rain".

O={S3,S3,S1} – observation sequence

P(O|Model)=P(S3,S3,S1|Model)=P(S3)*P(S3|S3)*P(S3|S1)=1*0.8*0.3=0.24

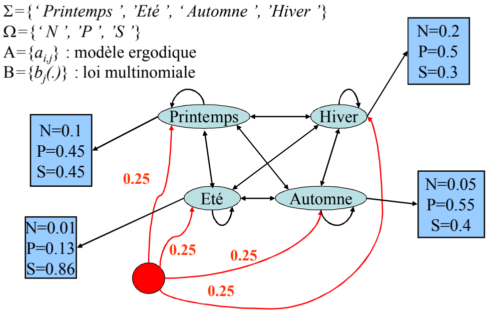
Un exemple de HMM

$\Sigma=\{$ '*Printemps* ', '*Eté* ', '*Automne* ', '*Hiver* '$\}$

$\Omega=\{$ '*N* ', '*P* ', '*S* '$\}$

$A=\{a_{i,j}\}$: modèle ergodique

$B=\{b_j(.)\}$: loi multinomiale



Que peut-on faire avec un HMM ?

- **Évaluation de modèle** : Étant donnés une séquence observée $O=o_1,\ldots,o_n$ et un HMM $\Lambda=\{\Sigma,\Omega,\Pi,A,B\}$, quelle est la vraisemblance de O suivant Λ, $P(O|\Lambda)$?
  - Calcul direct
  - Algorithme Forward-Backward
- **Explication** : Étant donnés un HMM Λ et une séquence observée O, quelle est la séquence d'états qui a la probabilité maximale d'avoir générée O ?
  - L'algorithme de Viterbi
- **Modélisation** (Apprentissage) : Partant d'un ensemble d'observations $O=\{O^1,\ldots,O^T\}$, comment régler les paramètres d'un HMM Λ pour maximiser la vraisemblance de $P(O|\Lambda)$?
  - L'algorithme de Baum-Welch

Quelques jalons historiques

	60
• HMM [Baum70]	70
• Traitement de la parole [Rab89]	….
• Reconnaissance de textes manuscrits [KHB88]	80
• Analyse de séquences biologiques [HKMS92] [DEKM98]	90
• Modélisation de signaux acoustiques [Raph98]	…
• Test de circuits micro-électroniques [BGC00]	00

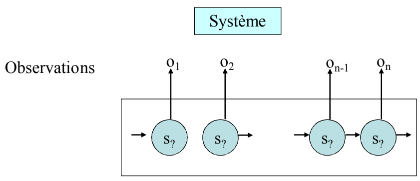
Quelques domaines d'application

- Reconnaissance de formes
- Reconnaissance de la parole
- Identification et contrôle
- Traitement du signal
- Analyse des séquences biologiques
- Économie
- Analyse géopolitique
- Robotique
- Diagnostic

Évaluation de modèle : L'algorithme forward-backward

- Étant donnés $\Lambda=\{\Sigma,\Omega,\Pi,A,B\}$ et une séquence observée $O=o_1,\ldots,o_n$, quelle est la vraisemblance de O suivant Λ, $P(O|\Lambda)$?

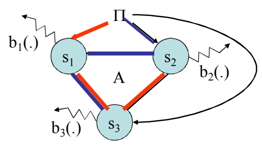
Explication : L'algorithme de Viterbi



On veut trouver la séquence d'états $I=s_{i(1)},\ldots,s_{i(n)}$ qui a la probabilité maximale d'avoir généré $O=o_1,\ldots,o_n$?

$\boxed{\max_I P(I|O,\Lambda)}$ ou, de manière équivalente : $\boxed{\max_I P(O,I|\Lambda)}$

L'algorithme de Viterbi (suite)



$\boxed{\max_I P(O,I|\Lambda)}$

$\Updownarrow$

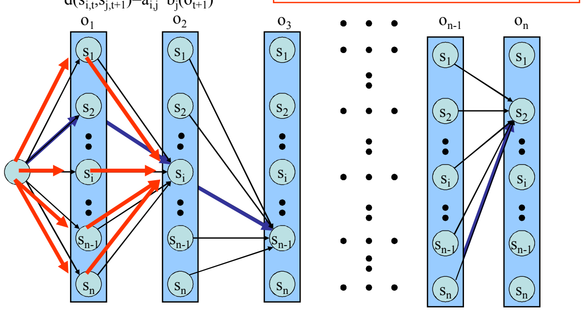
Recherche parmi tous les chemins possibles (m^n)

- Algorithme de Viterbi : programmation dynamique

Algorithme de Viterbi(suite)

De gauche à droite avec $d(s_{i,t},s_{j,t+1})=a_{i,j}*b_j(o_{t+1})$

Principe d'optimalité de Bellman



Apprentissage d'un HMM

A partir d'un ensemble d'observations $O=\{O^1,\ldots,O^T\}$, comment ajuster les paramètres de $\Lambda=<S,\Omega,\Pi,A,B>$ pour maximiser la vraisemblance de l'ensemble d'apprentissage, $P(O|\Lambda)$?

Problématiques

- Choix du nombre d'états (fixé, automatique (critères globaux, fusions d'états))
- Choix de la fonction d'émission (loi multinomiale, normale, Student)
- Méthodes d'apprentissage (Viterbi, Baum-Welch, NN)

Entraînement du HMM

On dispose d' un ensemble d'observations $O=\{O^1,\ldots,O^T\}$.

•Principe du max. de vraisemblance: $P(O|\Lambda)=\prod_{i=1}^{T} P(O^i|\Lambda)$

• Principe du max. de vraisemblance suivant les chemins de Viterbi:

$$P(O|\Lambda,V)=\prod_{i=1}^{T} P(O^i|\Lambda,V^i)$$

Etapes de Réalisation d'un système RAP et RAL de mots isolés sous HTK

- Enregistrement de la base et Étiquetage
- Paramètrisation: paramétres acoustiques
- Desscription des unités HMM de base du dictionnaire.

Algorithme Baum Welch: Hrest, Herest

Algorithme Viterbi: Hvite,

Toolkit HTK

Dictionnaire

- oh oh
- zero zero
- one one
- two two
- three three
- four four
- five five
- six six
- seven seven
- eight eight
- nine nine

Grammaire

- $chiffre = cinq|six|sept|huit|neuf;
- $pause=pause;
- $numerote=$chiffre [$pause] $chiffre [$pause] $chiffre [$pause] $chiffre [$pause] $chiffre [$pause];
- (SENT-START
- ($pause(numeroter)[$pause][$numerote][$pause])
- SENT-END)
- Construire l'HMM de chaque mot à reconnaitre (état et probabilité de transition)

Référence

- Adapté de Yannis Korilis, Christian St-Jean
- HTK Boite à outils