

modélisation de séquence pour l'interaction homme machine

-INSA de Rouen- ASI4 - IHME -

Clément Chatelain
clement.chatelain@insa-rouen.fr

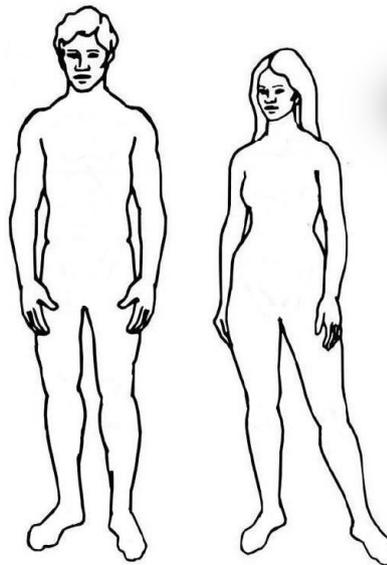
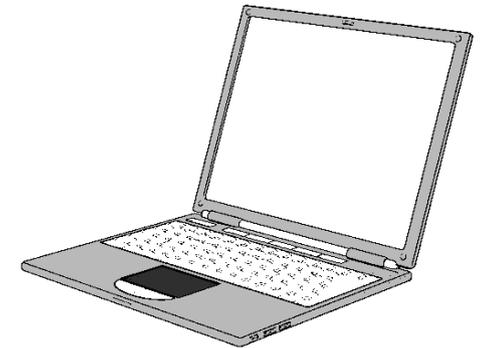
L'image du titre a été générée automatiquement grâce à :
<http://www.cs.toronto.edu/~graves/handwriting.html>

Plan

- Introduction à la modélisation de séquences
- Markov Model et Hidden Markov Model
- Application à la reconnaissance de la parole

Introduction

Interactions Homme - machine



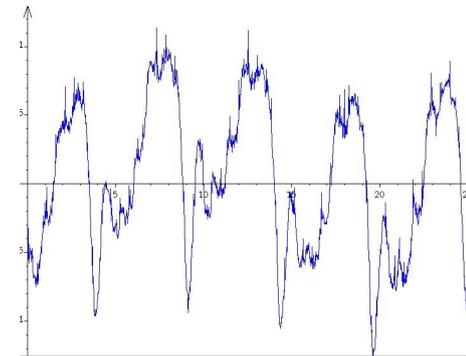
Et aussi : **Concatenation (grande fabrique)**



Introduction

- Quel est le point commun à tous ces signaux ?
 - Une séquence de caractères saisie au clavier
 - Voix
 - Écriture
 - Coordonnées des doigts sur une tablette
 - Mouvements d'un utilisateur
 - Signaux de la spherio
 - Séquence d'une manette 'hbddgbABAB'
 - Etc.

*Je vas pin de voir en
mon souvenir le meilleur,*



=> Leur Nature est séquentielle, dynamique

Modélisation de Séquence

- C'est quoi modéliser ?

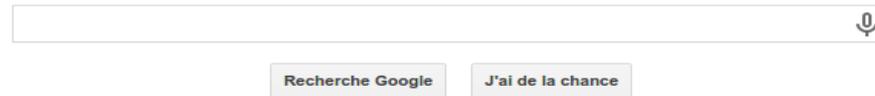
“élaborer un modèle, représenter un phénomène à l'aide d'un système qui possède des propriétés analogues à ce phénomène”

- Pour quoi faire ?

- Pour prédire : météo, bourse, etc.
- Pour reconnaître : parole, écriture, etc.
- Pour Simuler : médecine, biologie
- Pour faire des **interactions homme machine**

Exemples & démo

- Reconnaissance de la parole, par Google



- Génération d'écriture, par Alex Graves

modélisation de séquence pour l'interaction homme machine

Comment modéliser ?

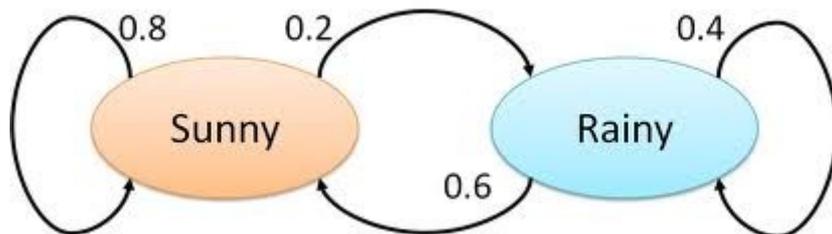
- Prendre en compte l'aspect séquentiel des signaux
 - Taille variable
 - Dépendances temporelles (exemple)
- Prendre en compte leur variabilité (intrinsèque + bruit)
 - Écriture/parole/etc.
- Prendre en compte des connaissances a priori
 - Dictionnaires, modèles de langage, etc.

"j ' ai malencontreusement marché sur le chion de ma vosrre . cele - ci , tenant brauemryp à son chien , a éle parthalièrement afestée par l ' acaident et me demande un dédammagement de taille "

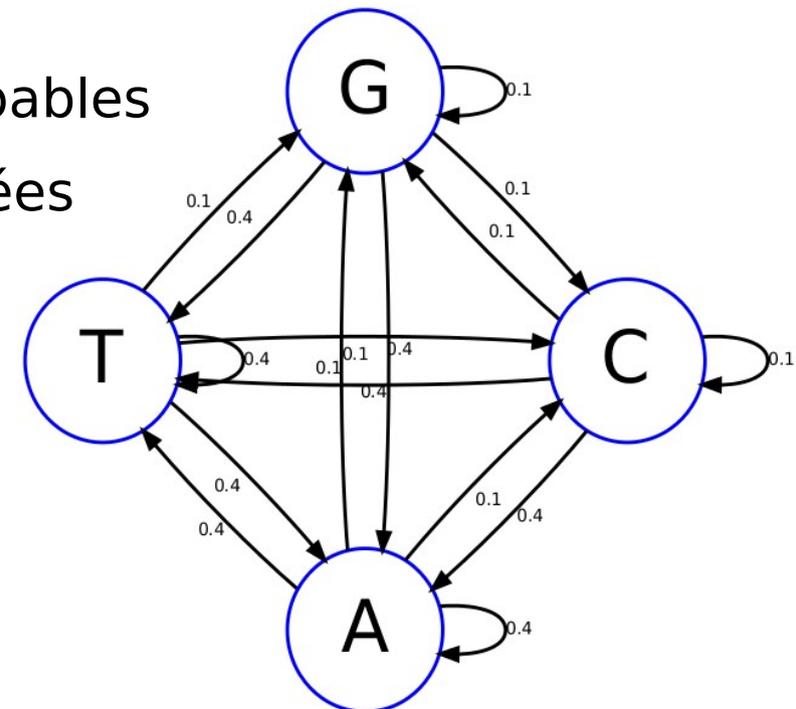
Solutions : modèle manuel < règles < apprentissage statistique

Markov Model

- Modélisation stochastique permettant de modéliser une séquence par une succession d'états
 - États discrets $x_i \in E$
 - Hypothèse markovienne : un état ne dépend que de l'état précédent
 - Les états sont plus ou moins probables
 - Transitions entre états probabilisées



$E = \{\text{Sunny}, \text{Rainy}\}$



$E = \{A, C, T, G\}$

Markov Model

- OK pour modéliser une séquence prise au clavier
- En revanche, dans le cas de l'interaction Homme machine, les états ne sont pas toujours observables. Exemple :
 - La machine ne voit pas ' **m o t** ' mais
 - La séquence d'états est donc cachée
- On va donc chercher à modéliser la séquence d'états et la manière dont elle génère des observations



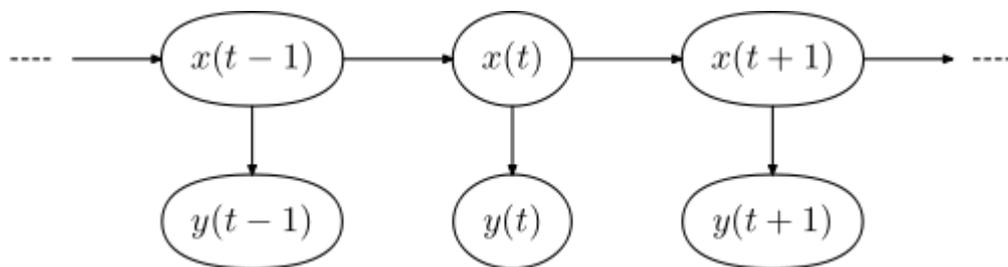
=> Hidden Markov Model

Hidden Markov Model (HMM)

Méthode la plus répandue depuis les années 80 pour la modélisation statistique de signaux.

Rabiner, L., "A tutorial on hidden Markov models and selected applications in speech recognition," Proceedings of the IEEE , vol.77, no.2, pp.257,286, Feb 1989

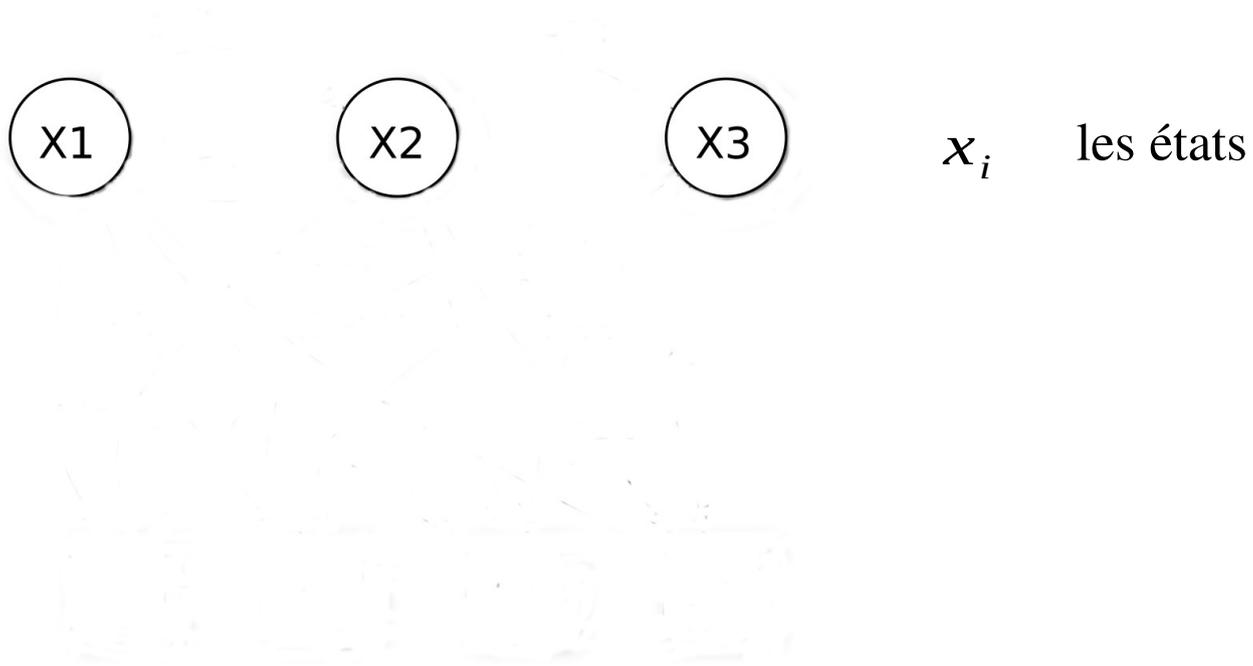
- Un HMM modélise le processus de génération d'une séquence d'observations par une chaîne de Markov.



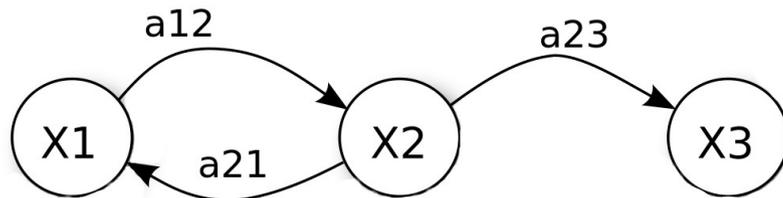
États cachés (= ' m o t ')

Observations (=  ou )

Définition d'un HMM



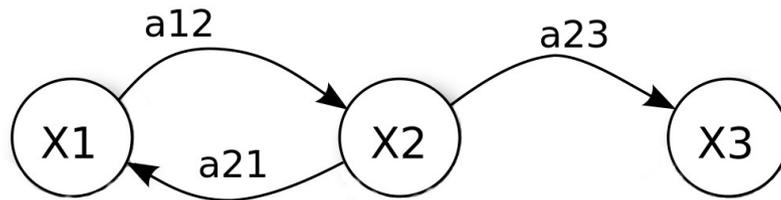
Définition d'un HMM



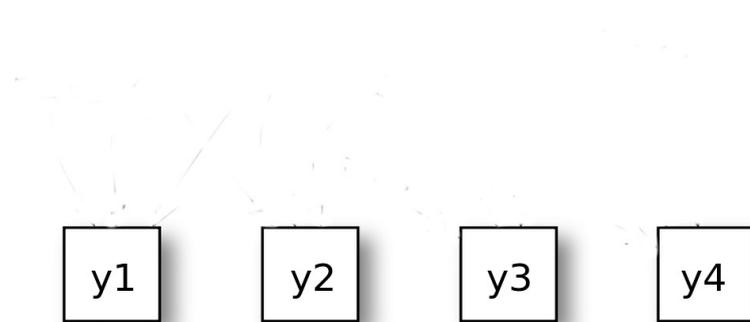
x_i les états (N états possibles)

$A = \{a_{ij}\}$ Matrice des probabilités de transition entre états

Définition d'un HMM



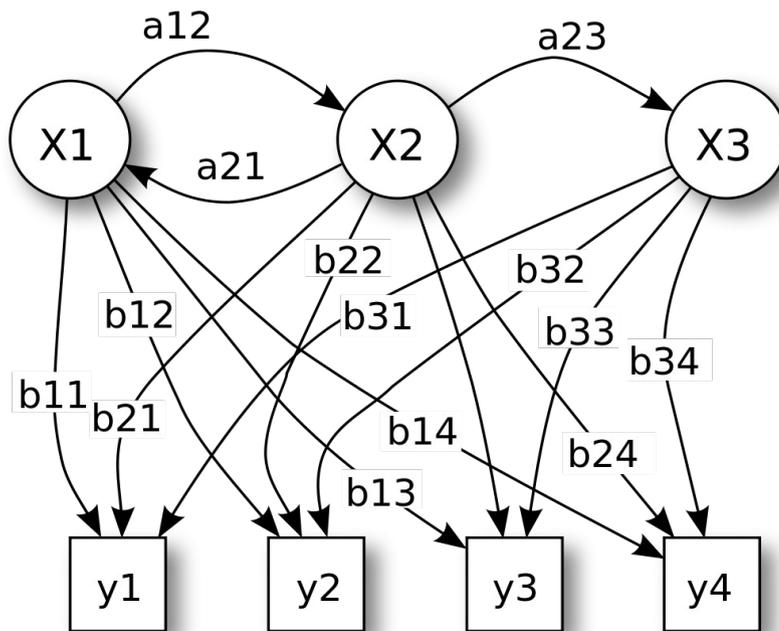
x_i les états (N états possibles)



y_j (ou o_j) les observations

$A = \{a_{ij}\}$ Matrice des probabilités de transition entre états

Définition d'un HMM



x_i les états (N états possibles)

y_j (ou o_j) les observations

$A = \{a_{ij}\}$ Matrice des probabilités de transition (de l'état i vers j)

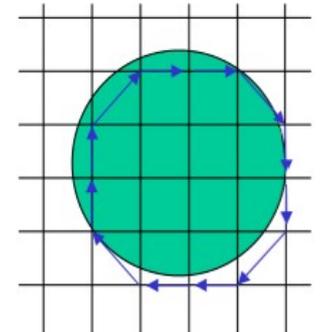
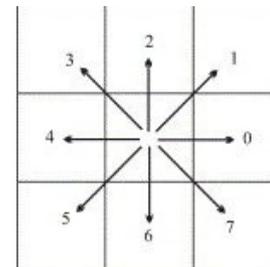
$B = \{b_{ji}\}$ ou $\{b_j(i)\}$ Matrice des probabilités d'émission (proba d'observer i en j)

$\Pi = \{\pi_i\}$ Distribution des états initiaux

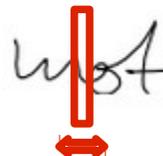
HMM discret/continu

- Si les observations y_j sont discrètes, le HMM est dit discret.
 - Alphabet fini de M symboles
 - Exemples :

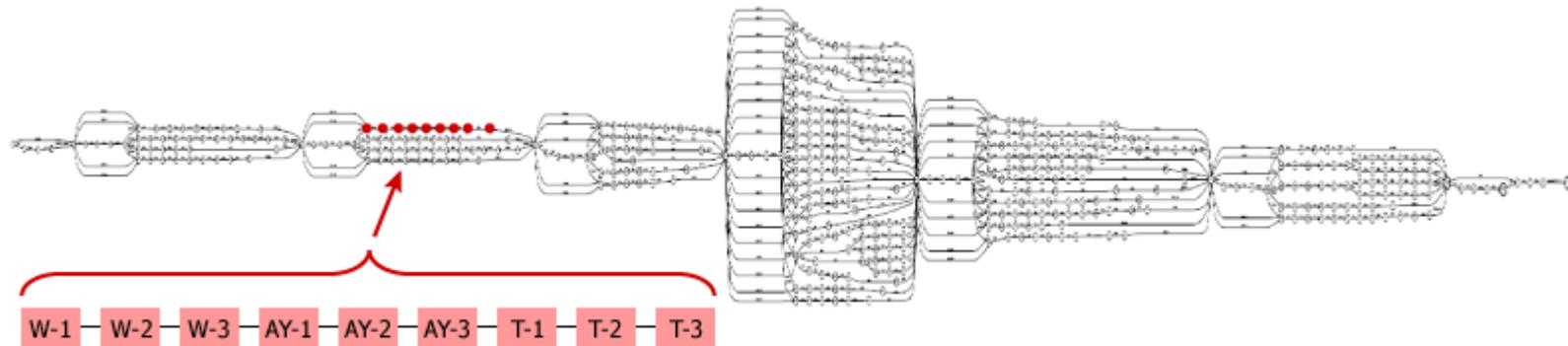
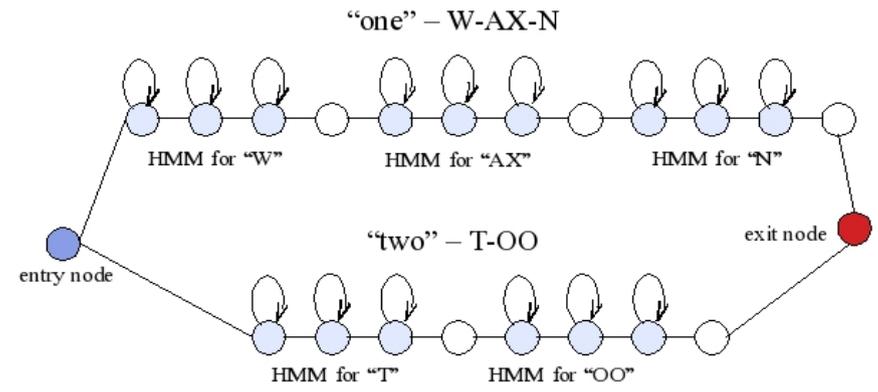
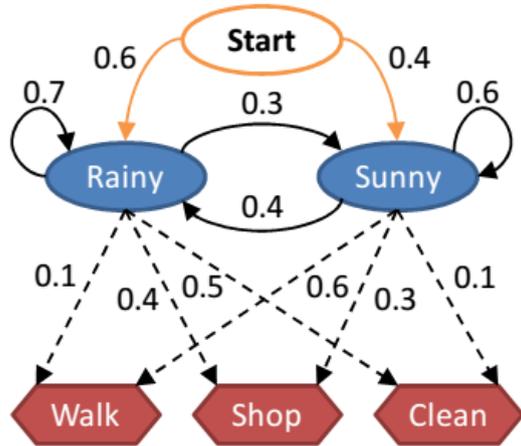
$x_i \in \{beau, pluie, nuageux\}; y_j \in \{parapluie, mouillé, Kway, lunettes\ de\ soleil\}$
 $x_i \in \{rond, carré, losange\}; y_j \in \{0,1,2,3,4,5,6,7\}$



- Si les observations sont continues, le HMM est dit continu.
 - Valeurs dans \mathbb{R} ou \mathbb{R}^p
 - Exemples :



Exemples de HMM



Les trois problèmes du HMM

- **Problème 1 (évaluation) :**

Etant donné une séquence d'observations $O=(o_1\dots o_T)$ et un modèle $\lambda=(A,B,\Pi)$, comment calculer efficacement $P(O/\lambda)$?

- **Problème 2 (décodage) :**

étant donné un modèle $\lambda=(A,B,\Pi)$, quelle est la séquence d'états $X=(x_1\dots x_n)$ la plus probable ayant pu générer la séquence d'observations $O=(o_1\dots o_T)$

- **Problème 3 (apprentissage) :**

Etant donnée une séquence d'observations $O=(o_1\dots o_T)$ comment ajuster les paramètres du modèle pour maximiser $P(O/\lambda)$?

Problème 1 : évaluation

Etant donné une séquence d'observations $O = (o_1 \dots o_T)$ et un modèle $\lambda = (A, B, \Pi)$, comment calculer efficacement $P(O|\lambda)$?

- **Algorithme Forward**

Solution exacte

$O(N^2T)$

The Forward-Backward Procedure [2], [3]⁶: Consider the forward variable $\alpha_t(i)$ defined as

$$\alpha_t(i) = P(O_1 O_2 \cdots O_t, q_t = S_i | \lambda) \quad (18)$$

i.e., the probability of the partial observation sequence, $O_1 O_2 \cdots O_t$, (until time t) and state S_i at time t , given the model λ . We can solve for $\alpha_t(i)$ inductively, as follows:

1) Initialization:

$$\alpha_1(i) = \pi_i b_i(O_1), \quad 1 \leq i \leq N. \quad (19)$$

2) Induction:

$$\alpha_{t+1}(j) = \left[\sum_{i=1}^N \alpha_t(i) a_{ij} \right] b_j(O_{t+1}), \quad 1 \leq t \leq T - 1$$
$$1 \leq j \leq N. \quad (20)$$

3) Termination:

$$P(O|\lambda) = \sum_{i=1}^N \alpha_T(i). \quad (21)$$

Problème 2 : décodage

étant donné un modèle Λ ,
quelle est la séquence d'états
 $X = (x_1 \dots x_n)$ la plus probable
ayant pu générer la séquence
d'observations $O = (o_1 \dots o_T)$?

- Algorithme de viterbi
Solution approchée

1) Initialization:

$$\delta_1(i) = \pi_i b_i(O_1), \quad 1 \leq i \leq N \quad (32a)$$

$$\psi_1(i) = 0. \quad (32b)$$

2) Recursion:

$$\delta_t(j) = \max_{1 \leq i \leq N} [\delta_{t-1}(i) a_{ij}] b_j(O_t), \quad 2 \leq t \leq T$$
$$1 \leq j \leq N \quad (33a)$$

$$\psi_t(j) = \operatorname{argmax}_{1 \leq i \leq N} [\delta_{t-1}(i) a_{ij}], \quad 2 \leq t \leq T$$
$$1 \leq j \leq N. \quad (33b)$$

3) Termination:

$$P^* = \max_{1 \leq i \leq N} [\delta_T(i)] \quad (34a)$$

$$q_T^* = \operatorname{argmax}_{1 \leq i \leq N} [\delta_T(i)]. \quad (34b)$$

4) Path (state sequence) backtracking:

$$q_t^* = \psi_{t+1}(q_{t+1}^*), \quad t = T - 1, T - 2, \dots, 1. \quad (35)$$

Problème 3 : apprentissage

étant donné un modèle Λ ,
quelle est la séquence d'états
 $X = (x_1 \dots x_n)$ la plus probable
ayant pu générer la séquence
d'observations $O = (o_1 \dots o_T)$?

Problème le plus difficile

- Pas de solution analytique
- Algorithme de Baum Welsh

Algorithme itératif
→ Solution approchée

Réestimation successive
Des paramètres du modèle :

$$\pi_i = \frac{\pi_i \frac{\partial P}{\partial \pi_i}}{\sum_{k=1}^N \pi_k \frac{\partial P}{\partial \pi_k}}$$

$$a_{ij} = \frac{a_{ij} \frac{\partial P}{\partial a_{ij}}}{\sum_{k=1}^N a_{ik} \frac{\partial P}{\partial a_{ik}}}$$

$$b_j(k) = \frac{b_j(k) \frac{\partial P}{\partial b_j(k)}}{\sum_{\ell=1}^M b_j(\ell) \frac{\partial P}{\partial b_j(\ell)}}$$

avec $P = P(O/\Lambda)$

Retour sur les HMM continus

- Si HMM discret, les o_t sont discrets, donc les b_{ij} sont des probabilités.
- Si HMM continu, $o_t \in \mathbb{R}^p$, il faut donc estimer une densité de probabilité à partir d'un vecteur de réels.
 - Utilisation de mélanges de gaussiennes :

$$b_j(o) = \sum_{i=1}^M c_{jm} G(o, \mu_{jm}, \sigma_{jm})$$

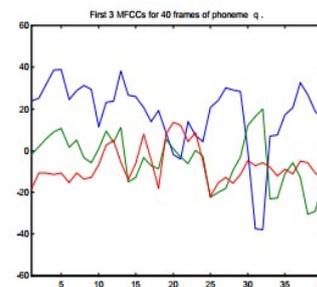
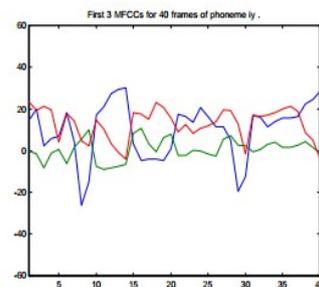
avec

$$\begin{cases} c_{jm} & \text{Coefficient de la } m^{\text{ème}} \text{ gaussienne} \\ G(o, \mu_{jm}, \sigma_{jm}) & \text{Gaussienne de vecteur moyen } \mu_{jm} \text{ et de matrice de covariance } \sigma_{jm} \end{cases}$$

- Les coefficients c_{jm} , μ_{jm} et σ_{jm} doivent être appris [Rabiner]
- Les $b_j(o)$ peuvent aussi être estimés par un classifieur (RdN)

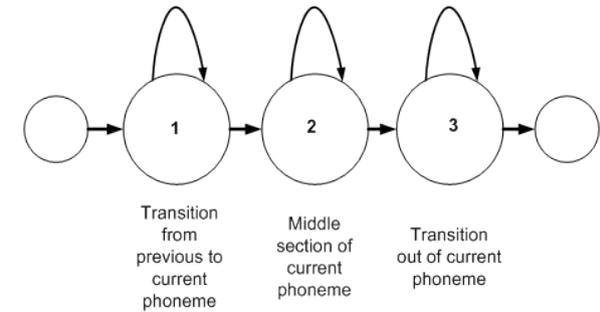
Reconnaissance de la parole

- Problème difficile : variabilité, bruit, lexique important (~100K mots)
- Utilisation de Fenêtre glissante + MFCC (Mel Frequency ceptral coeff.)
- Extraction de carac (energie, moyenne, etc.) → HMM continus



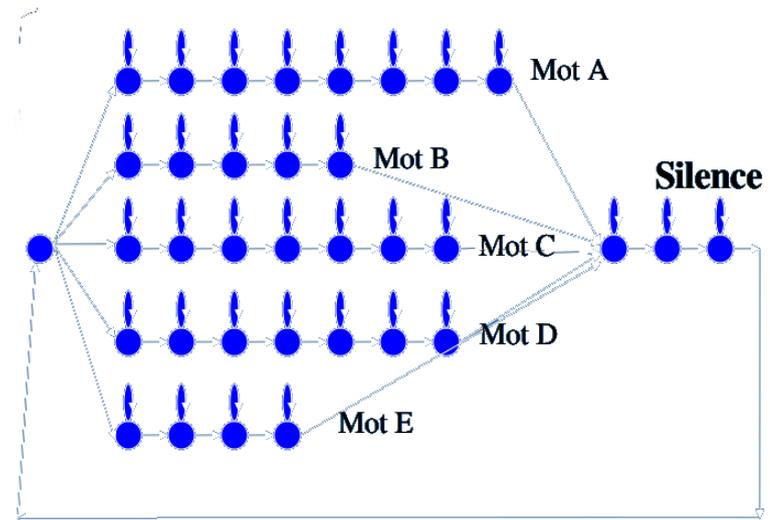
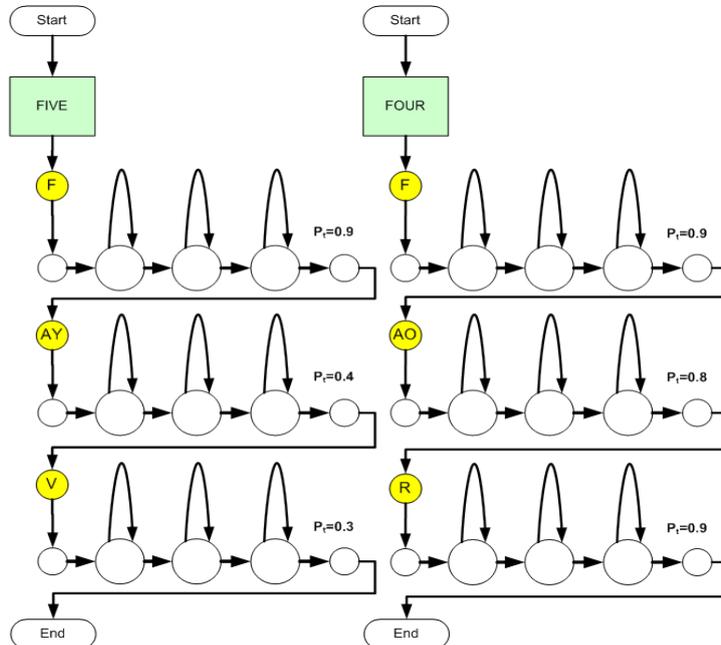
➔ Généralement qq dizaines de features

Reconnaissance de la parole



- **Les modèles :**

- Modélisation des phonèmes : qq états/phonèmes
- Modélisation des mots : concaténation de phonèmes
- Modélisation des phrases : concaténation des mots



Reconnaissance de la parole

- Ajout de modèles de langages (\neq modèles acoustiques)

- Réestimer les probabilités des hypothèses de reconnaissance

$$W = \operatorname{argmax}_w (P(O/W) p(W))$$

- Au niveau phonèmes : bigrammes, trigrammes

$$p('char_i', 'char_j')$$

- Au niveau mots : $p("mot_i", "mot_j")$

- Problèmes d'estimation

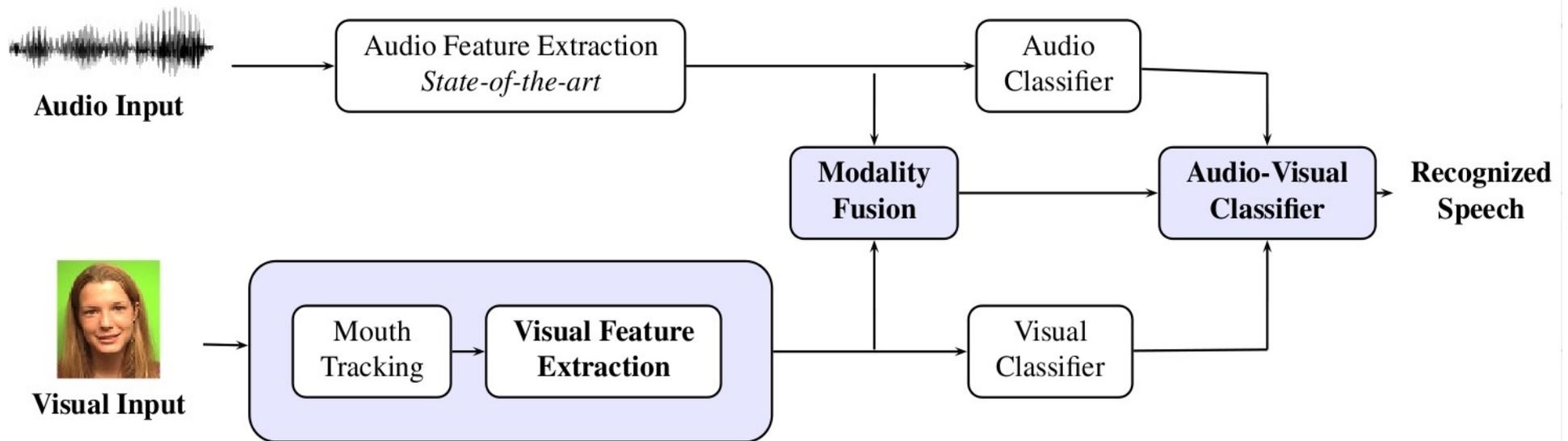
- 10K mots \rightarrow 1 Milliard de trigrammes de mots possibles ...
- \rightarrow entraînement sur de très grandes bases + Modèles de replis : estimations à partir de séquences plus courtes

- Problèmes de décodage

- Combinaison des infos acoustiques + langage
- Complexité du parcours des solutions

Reconnaissance de la parole

- Reconnaissance audio + video
 - Permet de traiter des occultations momentanées du son ou de la vidéo



Conclusion

- Autres applications des HMM en TAL (Traitement Automatique des Langues)
 - Traduction automatique
 - Génération de texte : brevets, bulletins météo

« THE HEAD AND IN FRONTAL ATTACK ON AN ENGLISH WRITER THAT THE CHARACTER OF THIS POINT IS THEREFORE ANOTHER METHOD FOR THE LETTERS THAT THE TIME OF WHO EVER TOLD THE PROBLEM FOR AN UNEXPECTED ».

[Shannon, ordre 2]

Conclusion

- Autres applications des HMM en TAL (Traitement Automatique des Langues)
 - Détection d'entités nommées
 - Correction d'orthographe
 - ...



Characters, bi-grams, or tri-gram of characters? Until this point, we only analysed bi-gram of characters in the distribution based stages. We see that bi-gram profiles allow great performances on these two datasets but we also evaluate the performances using the character profiles against the bi-gram of characters profiles and against tri-gram profiles. Character profiles might be more robust on difficult documents than bi-gram ones because of the OCR output reliability. Indeed, it is more difficult to get stability in the accuracy of having two consecutive characters. However, bi-gram profiles encode more knowledge and might be better for good quality documents. To evaluate the limits of knowledge encoding, we test both systems with tri-gram profiles. The results given in Table 8 are obtained with the weighted χ^2 comparison distance.

Un-Characters, bi-grams, or tri-gram of characters? Until this point, we only analysed bi-gram of characters in the distribution based stages. We see that bi-gram profiles allow great performances on these two datasets but we also evaluate the performances using the character profiles against the bi-gram of characters profiles and against tri-gram profiles. Character profiles might be more robust on difficult documents than bi-gram ones because of the OCR output reliability. Indeed, it is more difficult to get stability in the accuracy of having two consecutive characters. However, bi-gram profiles encode more knowledge and might be better for good quality documents. To evaluate the limits of knowledge encoding, we test both systems with tri-gram profiles. The results given in Table 8 are obtained with the weighted χ^2 comparison distance.

Personnage ONU, bi-grammes, ou tri-gramme de caractères? jusqu'à ce point, nous ne analysé bi-gramme de caractères dans les étapes de distribution fondés. Nous voyons que les profils bi-grammes permettent de grandes performances sur ces deux ensembles de données, mais nous évaluons également les performances en utilisant les profils ombre de ACTER contre la bi-gramme de profils de personnages et aux profils tri-grammes. profils de caractères peuvent être plus robuste sur les documents difficiles que les bi-grammes en raison de la sortie fiabilité OCR. En effet, il est plus difficile d'obtenir de la stabilité dans la précision de disposer de deux caractères consécutifs. Cependant, les profils bi-gram codent plus de connaissances et peut-être mieux pour documents de bonne qualité. Pour évaluer les limites de l'encodage de la connaissance, nous testons les deux systèmes avec des profils trigramme Les résultats présentés dans le Tableau 8 sont obtenus avec des profils bi-gram les pondérée χ^2 comparaison distance.

Test avec goggles

Conclusion

- HMM dans la pratique
 - HTK pour l'apprentissage des modèles
<http://htk.eng.cam.ac.uk/>
 - Base de données ???
TIMIT : <https://catalog.ldc.upenn.edu/LDC93S1> (250\$)
 - Julius pour le décodage grand lexique (open source)
http://julius.sourceforge.jp/en_index.php
- Méthodes à l'état de l'art pour la modélisation de séquence
 - Apprentissage statistique
 - Bases de données étiquetées = \$\$\$