

Traitement Automatique des Langues
pour les
Systèmes d'Information

Désambiguïstation morpho-syntaxique

Damien Nouvel
Nathalie Friburger
Jean-Yves Antoine

Désambiguïssation morpho-syntaxique

Plan



- Processus stochastiques
- Modèle de Markov caché (HMM)
- Algorithme de Viterbi
- Etiquetage morpho-syntaxique

Désambiguïisation morpho-syntaxique

Plan



- Processus stochastiques
- Modèle de Markov caché (HMM)
- Algorithme de Viterbi
- Etiquetage morpho-syntaxique

Désambiguïssation morpho-syntaxique

Processus stochastiques



- **Processus aléatoire :**

- Espace probabilisé des événements : (Ω, A, P)
- Séquence, suite, série, ensemble ordonné
 - **Indexation** d'évènements dans $T = \mathbb{R}, \mathbb{R}^+, \mathbb{N}$
- Peut-être mesuré en :
 - **Continu** : variables aléatoires
 - **Discret** : variables aléatoires \approx états (symboles)

- **Stochastique :**

- Indexation dans le **temps** (instant, date, indice)
- Réalisation d'un processus stochastique :
 - Variable aléatoire : $Y(\omega, t)$
 - Notation sur \mathbb{N} : $\mathbf{Y}(\boldsymbol{\omega}, \mathbf{t}) = \mathbf{Y}_t(\boldsymbol{\omega}) = \mathbf{Y}_t$

Désambiguïisation morpho-syntaxique

Processus stochastiques

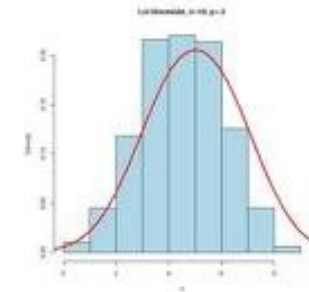
- Quelques processus stochastiques :

- Répétitions indépendantes d'une même expérience :

- Lancer « pile ou face » n fois :

- Avoir « pile » à l'instant t : p

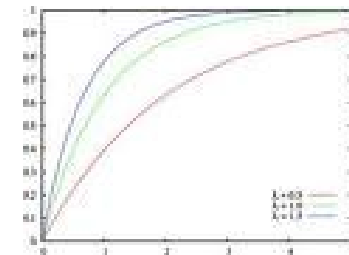
- Probabilité de k « piles » après n tirages (binomiale) : $C_n^k p^k q^{n-k}$



- Loi de probabilité dépendante du temps :

- Durée de vie d'un composant électronique :

- Probabilité qu'il soit HS au jour j : $1 - e^{-\lambda j}$



- Loi de probabilité dépendante des événements précédents :

- Tirer (sans remise) n fois dans un chapeau des tickets parmi T, dont g gagnants :

- Probabilité d'avoir k tickets gagnants : $C_g^k C_{T-g}^{n-k} / C_T^n$

- ...

Désambiguïssation morpho-syntaxique

Processus stochastiques



- **Processus stochastique**, une variable aléatoire :
 - Discrète ou continue
 - Plus ou moins **dépendante de paramètres (Ω)**
 - Plus ou moins **dépendante du temps (T)**
- Estimation :
 - Déterminer les **paramètres qui modélisent le mieux** le processus selon les dépendances et la loi de probabilité
- Prédiction :
 - Évaluer la **probabilité de la prochaine valeur** pour la variable étant donné les paramètres du modèle et l'historique :

$$P(Y_t = y \mid Y_{t-1} = y_{t-1}, Y_{t-2} = y_{t-2} \dots, Y_0 = y_0) ?$$

Désambiguïisation morpho-syntaxique

Plan



- Processus stochastiques
- Modèle de Markov caché (HMM)
- Algorithme de Viterbi
- Etiquetage morpho-syntaxique

Désambiguïisation morpho-syntaxique

Modèle de Markov caché (HMM)

- Processus **markovien** :
 - **Processus stochastique**
 - Une variable ne dépend que du **présent**, pas du passé
 - Utile en physique statistique, théorie de l'information...



- **Chaîne de Markov (1913)** :
 - Succession des lettres dans « Eugène Onéguine »
 - Temps discret, espace discret : $(Y_t = y \mid t \in \mathbb{N}, y \in \mathbb{N})$
 - La probabilité d'une variable ne dépend **que de celle la précédant dans le temps** :

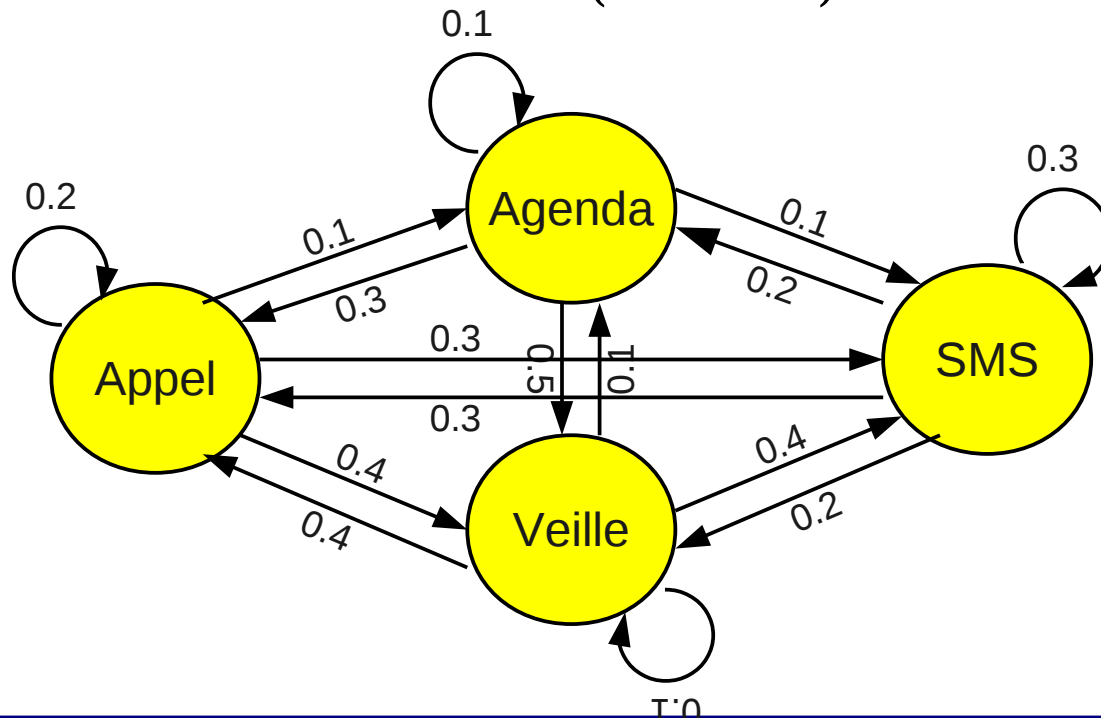
$$P(Y_t = y \mid Y_{t-1} = y_{t-1}, Y_{t-2} = y_{t-2} \dots Y_0 = y_0) = P(Y_t = y \mid Y_{t-1} = y_{t-1})$$

- **Pas de « mémoire »**, et « homogène » (ne dépend pas de t)

Désambiguïssation morpho-syntaxique

Modèle de Markov caché (HMM)

- Modélisation par chaîne de Markov :
 - Modèle « **graphique** »
 - **Variable** \approx **états**
 - Probabilités (matrice) de **transitions**, par ex. :

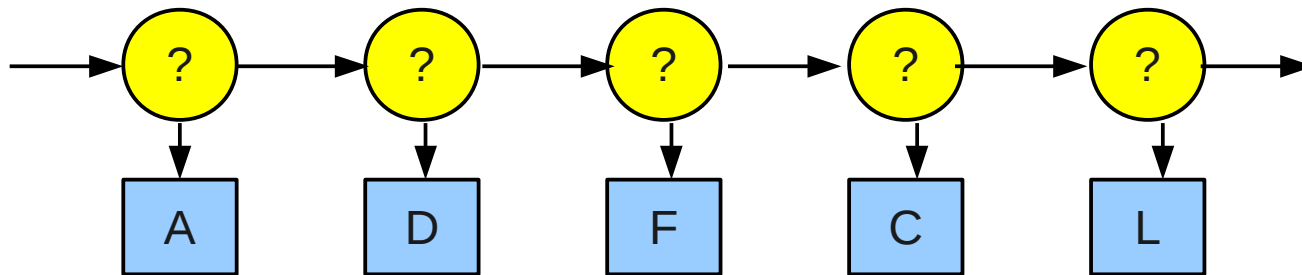


	Appel	Agenda	SMS	Veille
Appel	0.2	0.1	0.3	0.4
Agenda	0.3	0.1	0.1	0.5
SMS	0.3	0.2	0.3	0.2
Veille	0.4	0.1	0.4	0.1

Désambiguisation morpho-syntaxique

Modèle de Markov caché (HMM)

- Modèle de Markov **caché** :
 - **Hidden Markov Model (HMM) (Baum, 1965)**
 - Modélisation d'un processus stochastique « **génératif** » :
 - État du système : non connu
 - Connaissance pour chaque état des **probabilités** comme état initial, de **transition** entre états et de **génération** de symboles
 - **Observations** sur ce qu'a « généré » le système



- Applications : physique, reconnaissance de parole, traitement du langage, bio-informatique, finance, etc.

Désambiguïsation morpho-syntaxique

Modèle de Markov caché (HMM)



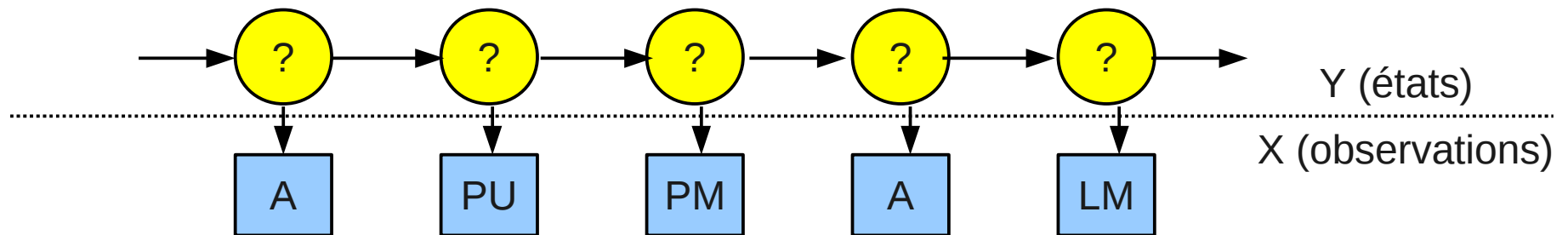
- Exemple sur le smartphone :
 - On ne connaît pas l'usage à un instant t du smartphone
 - Mais on observe que l'utilisateur réalise sur l'écran tactile du smartphone majoritairement l'un des contacts suivants :
 - Aucun contact (A)
 - Ponctuel unique (PU)
 - Ponctuel multitouch (PM)
 - Long unique (LU)
 - Long multitouch (LM)
- Enfin, on connaît la probabilité, dans chaque état, de réaliser chacun des contacts

	Appel	Agenda	SMS	Veille
A	0.1	0.1	0.1	0.5
PU	0.4	0.2	0.2	0.1
PM	0.1	0.2	0.5	0.1
LU	0.2	0.4	0.1	0.1
LM	0.2	0.1	0.1	0.2

Désambiguïisation morpho-syntactique

Modèle de Markov caché (HMM)

- Problématique :
 - **Étant donné une suite de symboles générés, quelle est la suite d'états (cachés) la plus probable ?**



- Par ex. :
 - $P(\text{Veille, Appel, SMS, Veille, Agenda}) ?$
 - $P(\text{Veille, SMS, SMS, Veille, Appel}) ?$
 - $P(\text{Veille, Appel, Veille, Veille, Veille}) ?$

Désambiguïssation morpho-syntaxique

Modèle de Markov caché (HMM)



- **Formalisation**, soit le HMM (A, B, π) tel que :

- Ensemble d'**états** : $\{ s_1, s_2 \dots s_n \}$

- Vecteur de **probabilité d'états initiaux** :

$$\pi = \{ p_i = p(Y_0 = s_i) \}$$

- Matrice de **probabilités de transitions** :

$$A = \{ a_{i,j} = p(Y_t = s_j \mid Y_{t-1} = s_i) \} \quad (\sum_i a_{i,j} = 1)$$

- Ensemble d'**observables** : $\{ o_1, o_2 \dots o_k \}$

- Matrice de **probabilités d'observations** :

$$B = \{ b_{i,l} = p(X_t = o_l \mid Y_t = s_i) \} \quad (\sum_i b_{i,l} = 1)$$

Désambiguïssation morpho-syntaxique

Modèle de Markov caché (HMM)



- Probabilité d'observations :
 - Indépendance des **transitions d'états** :

$$P(Y_1 = s_{t(1)}, Y_2 = y_{t(2)} \dots Y_n = s_{t(n)}) = \pi_{t(1)} * a_{t(1),t(2)} * \dots * a_{t(n-1),t(n)}$$

- Indépendance des **observations sachant les états** :

$$P(X_1 = o_{u(1)}, X_2 = o_{u(2)} \dots X_n = o_{u(n)} \mid Y_1 = s_{t(1)}, Y_2 = s_{t(2)} \dots Y_t = s_{t(n)}) \\ = b_{t(1),u(1)} * b_{t(2),u(2)} \dots b_{t(n),u(n)}$$

- On en déduit (**Bayes**) :

$$P(X_1 = o_{u(1)}, X_2 = o_{u(2)} \dots X_n = o_{u(n)}) \\ = \sum_{t(i)} \pi_{t(1)} * b_{t(1),u(1)} * a_{t(1),t(2)} * b_{t(2),u(2)} * a_{t(2),t(3)} \dots b_{t(n-1),u(n-1)} * a_{t(n-1),t(n)} * b_{t(n),u(n)}$$

- Estimation des paramètres du modèle à partir d'observation par l'algorithme de Viterbi / Forward Backward (Baum Welch)

Désambiguïssation morpho-syntaxique

Plan



- Processus stochastiques
- Modèle de Markov caché (HMM)
- Algorithme de Viterbi
- Etiquetage morpho-syntaxique

Désambiguïisation morpho-syntaxique

Algorithme de Viterbi



- Objectif : trouver la séquence d'états la plus probable étant donné un HMM et une séquence d'observation :

$$\max_{t()}\left[\mathbf{P}(Y_1 = \mathbf{s}_{t(1)} \dots Y_n = \mathbf{s}_{t(n)} \mid X_1 = \mathbf{o}_{u(1)} \dots X_n = \mathbf{o}_{u(n)})\right]$$

- Complexité :
 - Pour t observations, sur un HMM à s états, le nombre de séquences d'états possibles est s^t
 - Par ex. pour le smartphone, avec la séquence (A, PU, PM, A, LM), $4^5 = 1024$ solutions à examiner...
- Programmation dynamique :
 - Construction **incrémentale** de la solution la plus **probable**
 - Tire parti de la **chaîne de Markov**

Désambiguïisation morpho-syntaxique

Algorithme de Viterbi



- On **parcourt les observations** $(o_{u(1)}, o_{u(2)} \dots o_{u(n)})$ et on stocke pour chaque instant :

- La **meilleure probabilité d'atteindre chaque état** :

$$\alpha_0(i) = \pi_i * b_{i,u(1)}$$

$$\alpha_t(j) = \max_i [\alpha_{t-1}(i) * a_{i,j}] * b_{j,u(t)}$$

- L'état **précédent qui a donné la meilleure probabilité d'arriver à un état donné** :

$$\Psi_t(j) = \{ i \text{ t. q. } \alpha_{t-1}(i) * a_{i,j} = \max_k [\alpha_{t-1}(k) * a_{k,j}]$$

- Ainsi on ne conserve pas en mémoire les chemins qui ne conduisent pas aux meilleures probabilités
 - Possibilité de plus d'élaguer l'arbre de recherche

Désambiguisation morpho-syntaxique

Algorithme de Viterbi



- Par ex. :

- Observation A :

$$\alpha_0(\text{Appel}) = 0.25 * 0.1 = 0.025$$

$$\alpha_0(\text{Agenda}) = 0.25 * 0.1 = 0.025$$

$$\alpha_0(\text{SMS}) = 0.25 * 0.1 = 0.025$$

$$\alpha_0(\text{Veille}) = 0.25 * 0.5 = 0.125$$

- Observation PU :

$$\begin{aligned} \alpha_1(\text{Appel}) &= \max(0.025 * 0.2, 0.025 * 0.3, 0.025 * 0.3, 0.125 * 0.4) * 0.4 \\ &= 0.02 \quad (\Psi_1(\text{Appel}) = \text{Veille}) \end{aligned}$$

$$\begin{aligned} \alpha_1(\text{Agenda}) &= \max(0.025 * 0.1, 0.025 * 0.1, 0.025 * 0.2, 0.1 * 0.125) * 0.2 \\ &= 0.0025 \quad (\Psi_1(\text{Agenda}) = \text{Veille}) \end{aligned}$$

$$\begin{aligned} \alpha_1(\text{SMS}) &= \max(0.025 * 0.3, 0.025 * 0.1, 0.025 * 0.3, 0.125 * 0.4) * 0.2 \\ &= 0.01 \quad (\Psi_1(\text{Agenda}) = \text{Veille}) \end{aligned}$$

...

Désambiguïssation morpho-syntaxique

Algorithme de Viterbi



- Modélisation du problème :
 - **Processus stochastique**
 - **Observations** mais pas de connaissance des états
 - Connaissance des probabilités :
 - **Initiales**
 - **De Transitions**
 - **De génération de symboles**
 - Problème adapté pour un **modèle de Markov caché**
- Prédiction par l'**algorithme de Viterbi** :
 - Fournit la **solution optimale** pour les paramètres du modèle
 - **Sans fouiller tout l'espace de recherche** (prog. dynamique)

Désambiguïssation morpho-syntaxique

Plan



- Processus stochastiques
- Modèle de Markov caché (HMM)
- Algorithme de Viterbi
- Etiquetage morpho-syntaxique

Désambiguïssation morpho-syntaxique

Étiquetage morpho-syntaxique



- Catégorie morpho-syntaxique (MS) :
 - \approx **catégorie grammaticale** des mots
 - Déterminant
 - Nom
 - Verbe
 - Adjectif
 - Adverbe
 - Préposition
 - ...
 - \approx **rôle** du mot dans la phrase ?
 - Utile pour **analyser le langage** (chunking, syntaxe, RI, etc)
 - **Problème** : les mots sont **ambigus** (avions, brise, porte...)

Désambiguïisation morpho-syntaxique

Etiquetage morpho-syntaxique



- Par ex. : « La petite brise la glace »
 - DET ADJ NOM PRO VER
 - DET NOM VER DET NOM
- Dans la majorité des cas, la phrase n'est pas ambiguë, mais **certains mots** le sont :
 - Je porte le sac à dos ≠ Elle a fermé la porte.
 - Les avions décollent ≠ Nous **avions** beaucoup mangé.
 - ...
- Le « rôle » du token **détermine sa catégorie MS** :
 - Fortement dépendante de la catégorie MS précédente
 - Pas besoin de traitements sur le sens / la sémantique

Désambiguïssation morpho-syntaxique

Etiquetage morpho-syntaxique

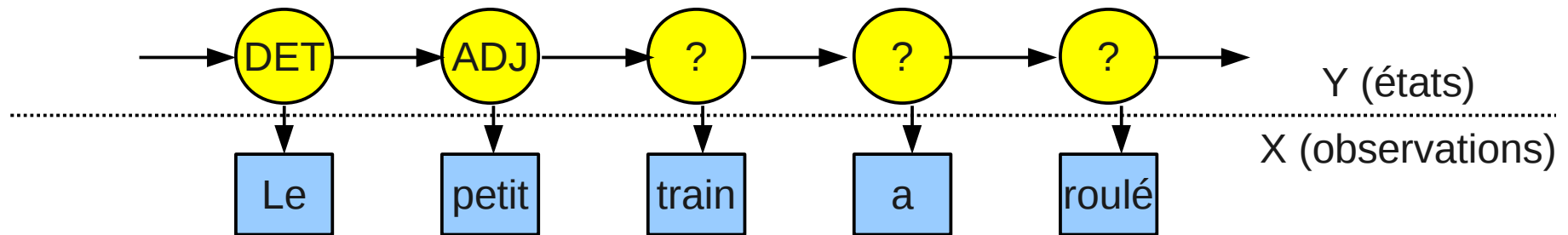


- Modèle de Markov bien adapté :
 - **Etats** : catégories MS (NOM, VER, PREP...)
 - **Probabilités initiales** : début par une catégorie MS :
 - $P(Y_0 = \text{DET}) = 0.3$
 - **Transitions** : probabilités que deux catégories MS se suivent
 - $P(Y_i = \text{NOM} \mid Y_{i-1} = \text{DET}) = 0.6$
 - $P(Y_i = \text{ADJ} \mid Y_{i-1} = \text{DET}) = 0.4$
 - $P(Y_i = \text{NOM} \mid Y_{i-1} = \text{ADJ}) = 0.8$
 - **Observations** : probabilités d'avoir un token par état :
 - $P(X_i = \text{avions} \mid Y_i = \text{NOM}) = 0.02$
 - $P(X_i = \text{porte} \mid Y_i = \text{VER}) = 0.01$

Désambiguïsation morpho-syntaxique

Étiquetage morpho-syntaxique

- Modèle « **génératif** » du langage pour la tâche d'étiquetage MS :



- Deux étapes :
 - **Apprentissage** du modèle par estimation des paramètres optimaux (Baum Welch) sur un corpus déjà étiqueté
 - **Prédiction** d'étiquetage sur d'autres textes (tokenisés) à l'aide du modèle paramétré et de l'algorithme de Viterbi