Automate Stochastique

Partie 3/3

Par *Guillaume Guérard*, Nouvelles Energies

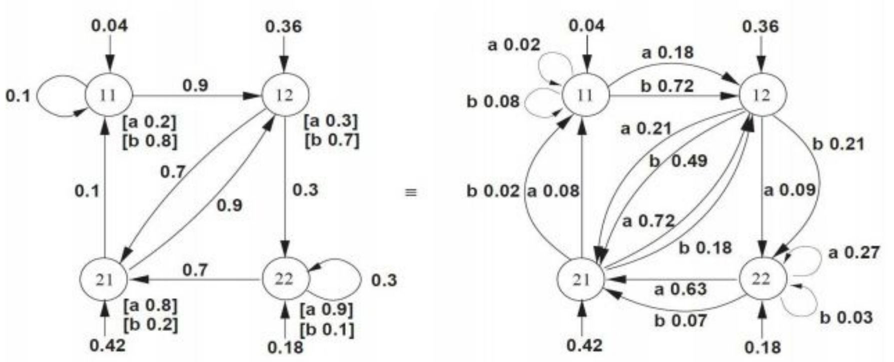
Equivalence

Hidden Markov Chain

Automate et Markov séparé

Probabilistic Finite Automata

On multiplie les probabilités d'observation avec les transitions cachées



Plus d'information sur <u>smart--grid.net</u>

L'ensemble des cours sont disponible au format Slides et Cours sur le site.

Equivalence

Hidden Markov Model

Les HMM sont des PFA où les symboles sont émis par les états et non les transitions.

Il n'existe pas de probabilité de "terminer" la chaîne d'observation.

Probabilistic Finite Automata

Les états possèdent des probabilités de départ, de fin et de transitions.

Le choix de la transition et l'observation sont généré par la même transition, simplifiant les calculs par rapport aux HMMs.

Equivalence

Probabilistic Finite Automata

Les états possèdent des probabilités de départ, de fin et de transitions.

Le choix de la transition et l'observation sont généré par la même transition, simplifiant les calculs par rapport aux HMMs.

Méthode d'apprentissage

Par ceux des HMM : Baum-Welch

Par inference Bayesienne (Gibbs)

Par réunion d'état - uniquement pour le cas déterministe (ALERGIA, Learn-PSA, MDI)

Plus d'information sur <u>smart--grid.net</u>

L'ensemble des cours sont disponible au format Slides et Cours sur le site.

Plan

1. PFA
2. HMM ⇔ PA
3. Merge & Fold
4. Mots interdits
5. Inférence

1 Automate probabiliste

Automate probabiliste

Un PFA est défini par $A = \{\Sigma, Q, \varphi, \iota, \tau\}$

- Σ un alphabet fini
- Q un nombre fini d'états
- $\varphi : Q \times \Sigma \times Q \rightarrow [0, 1]$ une fonction de transition
- $\iota : Q \rightarrow [0, 1]$ la distribution intiale
- $\tau : Q \rightarrow [0, 1]$ la probabilité d'arrêt

Automate probabiliste

Un PFA est défini par A = {Σ, Q, φ, ι, τ }

- Distribution initiale stochastique : $\sum_{q \in Q} \iota(q) = 1$
- Transition + Arrêt stochastique : $\forall q \in Q, \tau(q) + \sum_{a \in \Sigma} \sum_{q' \in Q} \phi(q, a, q') = 1$
- Un état initial : $\iota(q) > 0$
- Un état final : $\tau(q) > 0$
- Atteinte d'un état final : $P_{A_q}(\Sigma^*) = \sum_{q'} \phi(q, \Sigma^*, q')\tau(q') > 0.$

Automate probabiliste

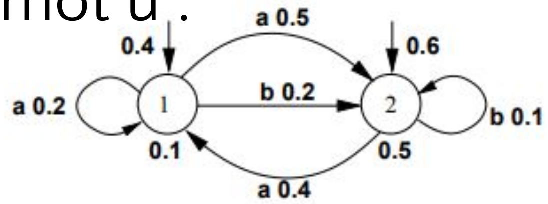
Un PFA est défini par A = {Σ, Q, φ, ι, τ }

- Probabilité de générer un prefix u :

$$\overline{P}_A(u) = \sum_{q,q' \in Q} \iota(q)\phi(q,u,q')$$

- Probabilité de générer un mot u :

$$P_A(u) = \sum_{q,q' \in Q} \iota(q)\phi(q,u,q')\tau(q')$$



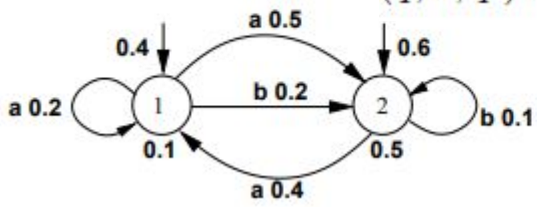
$$\begin{aligned} P_A(b) &= \iota(1)\phi(1,b,1)\tau(1) + \iota(1)\phi(1,b,2)\tau(2) \\ &+ \iota(2)\phi(2,b,1)\tau(1) + \iota(2)\phi(2,b,2)\tau(2) \\ &= 0.07 \end{aligned}$$

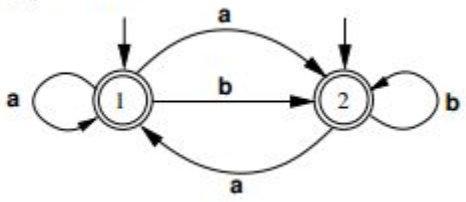
Automate probabiliste

Langage du PFA

- Support du PFA {Σ, Q, φ, ι, τ} = NFA {Σ, Q, δ, I, F}

$$(q, a, q') \in \delta \Leftrightarrow \phi(q, a, q') > 0$$





- Langage du PFA = langage du NFA

Automate probabiliste

Un DPFA est défini par A = {Σ, Q, φ, ι, τ }

- Un état initial unique

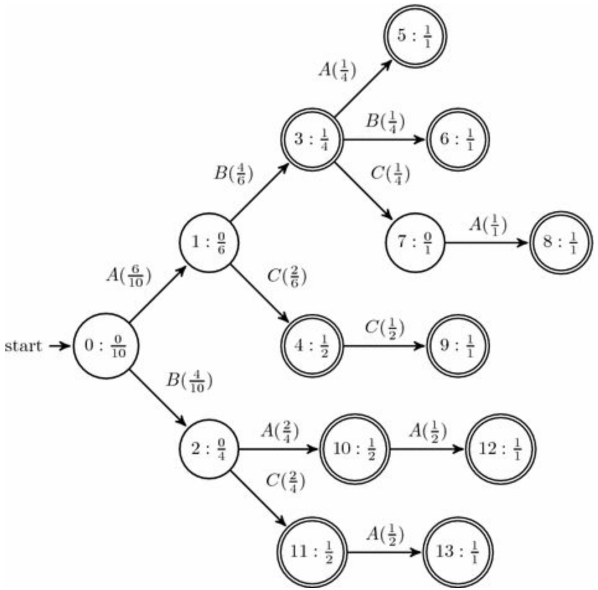
$$\exists q_0 \in Q \;;\; I(q_0) = 1$$

- Une transition par symbole unique

$$\forall q \in Q,\ \forall a \in \Sigma,\quad |\{q' : (q, a, q') \in \delta\}| \;\leq\; 1$$

Automate probabiliste

Un PPTA est DPFA sous forme d’arbre probabiliste ou de fréquences

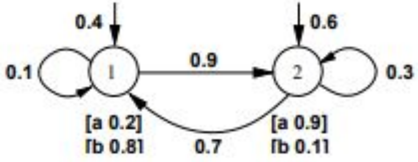


2

HMM <=> PA

HMM ⇔ PA

Rappel : HMM définie par M = {Σ, Q, A, B, ı}



- Σ un alphabet fini
- Q un ensemble fini d'états
- Probabilité de transition A : Q × Q → [0, 1]

$$\forall q \in Q, \sum_{q' \in Q} A(q, q') = 1$$

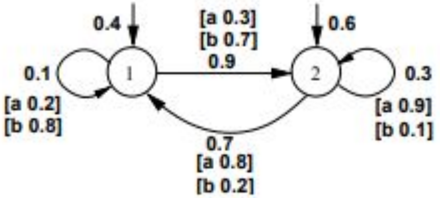
- Probabilité d'observation B : Q × Σ → [0, 1]

$$\forall q \in Q, \sum_{a \in \Sigma} B(q, a) = 1$$

- Probabilité initiale ı : Q → [0, 1] $\quad \sum_{q \in Q} \iota(q) = 1$

HMM ⇔ PA

HMMT définie par M = {Σ, Q, A, B, ı}



- Σ un alphabet fini
- Q un ensemble fini d'états
- Probabilité de transition A : Q × Q → [0, 1]

$$\forall q \in Q, \sum_{q' \in Q} A(q, q') = 1$$

- Probabilité d'observation B : Q × Σ × Q → [0, 1]

$$\boxed{\forall q, q' \in Q, \sum_{a \in \Sigma} B(q, a, q') = \begin{cases} 1 \text{ if } A(q, q') > 0 \\ 0 \text{ otherwise.} \end{cases}}$$

- Probabilité initiale ı : Q → [0, 1] $\quad \sum_{q \in Q} \iota(q) = 1$

HMM ⇔ PA

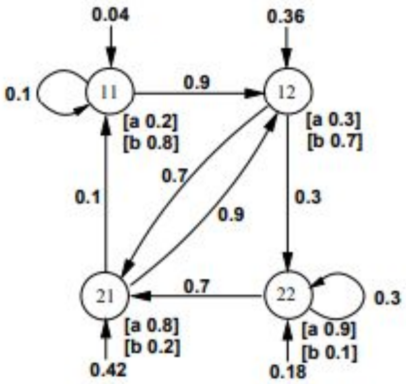
PFA et HMMT

$$A(q, q') = \sum_{a \in \Sigma} \phi(q, a, q')$$

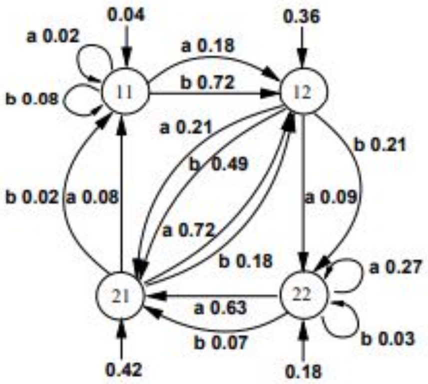
$$B(q, a, q') = \begin{cases} \dfrac{\phi(q,a,q')}{\sum_{a \in \Sigma} \phi(q,a,q')} & \text{if } \sum_{a \in \Sigma} \phi(q, a, q') > 0 \\ 0 & \text{otherwise.} \end{cases}$$

HMM ⇔ PA

HMM et PFA



⇒



$\phi(q, a, q') = B(q, a)A(q, q')$

$\forall q, \tau(q) = 0$

HMM ⇔ PA

HMM et HMMT version non-déterministe

$$Q' = \{(q, q') \in Q \times Q | A(q, q') > 0\}.$$

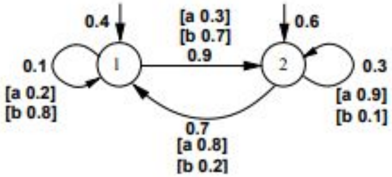
$$A((q, q'), (q'', q''')) = \begin{cases} A(q'', q''') & \text{if } q' = q'' \\ 0 & \text{otherwise.} \end{cases}$$

$$B((q, q'), a) = B(q, a, q')$$

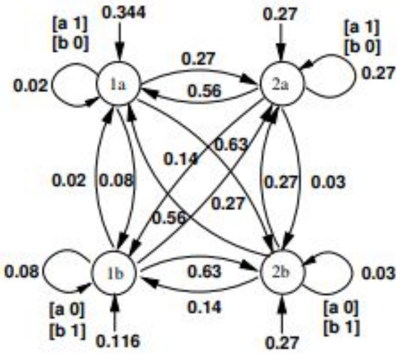
$$\iota'((q, q')) = \iota(q) A(q, q')$$

HMM ⇔ PA

HMM et HMMT version déterministe



⇒



$Q' = Q \times \Sigma \Rightarrow |Q'| = \mathcal{O}(|Q| \times |\Sigma|)$

$\iota'((q,a)) = \sum_{q' \in Q} \iota(q')A(q',q)B(q',a,q)$

$B'((q,a),x) = 1$ if $x = a$, and 0 otherwise

$A'((q,a),(q',b)) = A(q,q')B(q,b,q')$

HMM ⇔ PA

Probabilité d'émission d'une observation :
Backward-Forward

$$\Pr_{\mathcal{A}}(\theta) = I(s_0) \cdot \left(\prod_{j=1}^{k} P(s_{j-1}, x'_j, s_j) \right) \cdot F(s_k)$$

- Préfixe de taille i arrivant sur état q

$$\alpha_x(i, q) = \sum_{(s_0, s_1, \ldots, s_i) \in \Theta_{\mathcal{A}}(x_1 \ldots x_i)} I(s_0) \cdot \prod_{j=1}^{i} P(s_{j-1}, x_j, s_j) \cdot 1(q, s_i)$$

$1(q, q') = 1$ if $q = q'$ and 0 if $q \neq q'$.

$$\Pr_{\mathcal{A}}(x) = \sum_{q \in Q} \alpha_x(|x|, q) \cdot F(q)$$

$$\alpha_x(0, q) = I(q),$$

$$\alpha_x(i, q) = \sum_{q' \in Q} \alpha_x(i-1, q') \cdot P(q', x_i, q), \quad 1 \leq i \leq |x|$$

HMM ⇔ PA

Probabilité d'émission d'une observation : Backward-Forward

$$\Pr_{\mathcal{A}}(\theta) = I(s_0) \cdot \left(\prod_{j=1}^{k} P(s_{j-1}, x'_j, s_j) \right) \cdot F(s_k)$$

- Suffixe de taille i provenant de l'état q

$$\beta_x(i, q) = \sum_{(s_i, \ldots, s_{|x|}) \in \Theta_{\mathcal{A}}(x_{i+1} \cdots x_{|x|})} 1(q, s_i) \cdot \left(\prod_{j=i+1}^{|x|} P(s_{j-1}, x_j, s_j) \right) \cdot F(s_{|x|}),$$

$$\begin{aligned} \beta_x(|x|, q) &= F(q), \\ \beta_x(i, q) &= \sum_{q' \in Q} \beta_x(i+1, q') \cdot P(q, x_i, q'), \quad 0 \leq i \leq |x| - 1 \end{aligned}$$

$$\Pr_{\mathcal{A}}(x) = \sum_{q \in Q} I(q) \cdot \beta_x(0, q)$$

HMM ⇔ PA

Chemin générant l'observation avec la plus grande probabilité $\Pr_{\mathcal{A}}(\theta) = I(s_0) \cdot \left(\prod_{j=1}^{k} P(s_{j-1}, x'_j, s_j) \right) \cdot F(s_k)$ $\tilde{\theta} = \underset{\theta \in \Theta_{\mathcal{A}}(x)}{\operatorname{argmax}} \Pr_{\mathcal{A}}(\theta)$

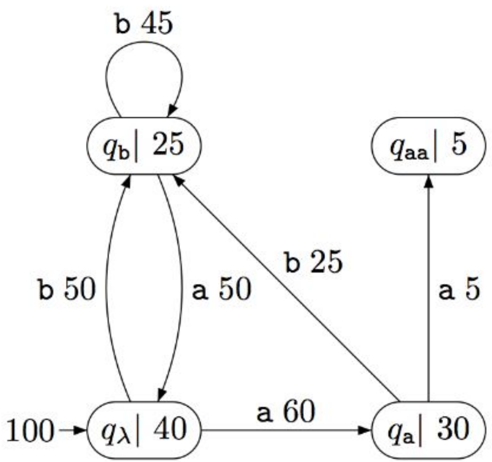
- Transformation en HMM
- Application de Viterbi

3

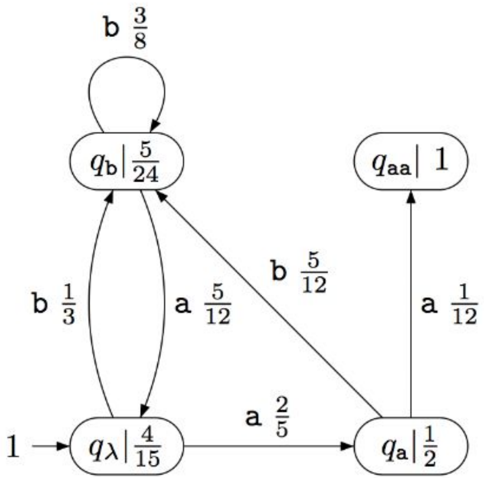
Merge & Fold

C'est cool les origami !!!

Merge & Fold



Frequency finite automata: FFA



Probabilistic finite automata: PFA

$$P(q) = \frac{F(q)}{F_{q' \in Q, a \in \Sigma}(q', a, q)}$$

Probability of states

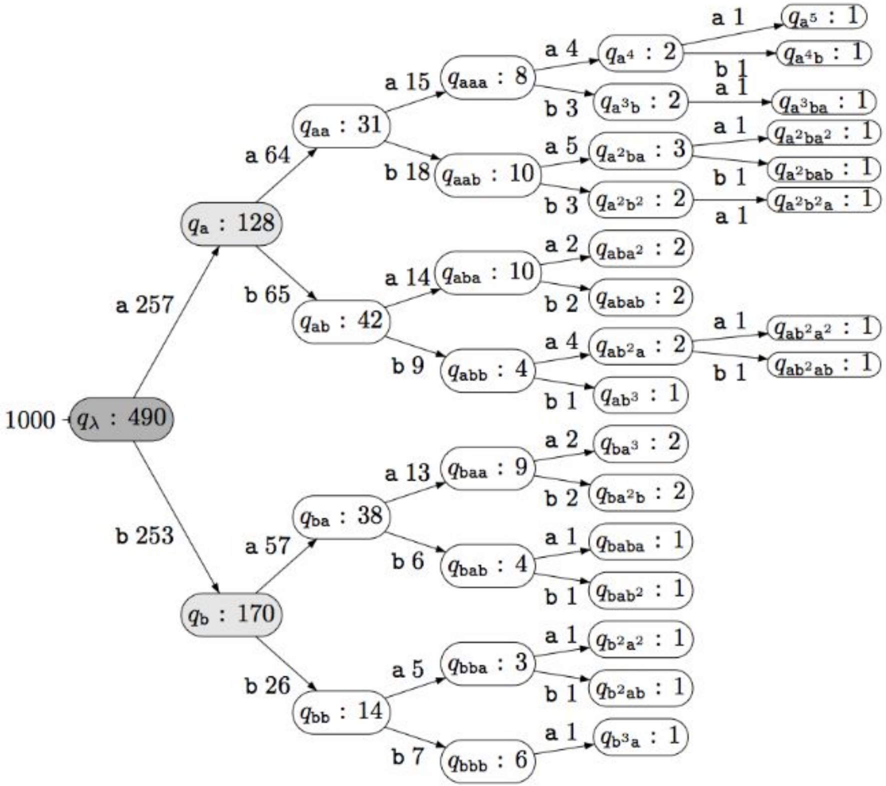
$$\partial(q, a, q') = \frac{F(q, a, q')}{F_{q'' \in Q, b \in \Sigma}(q'', b, q)}$$

Probability of edges

Merge & Fold

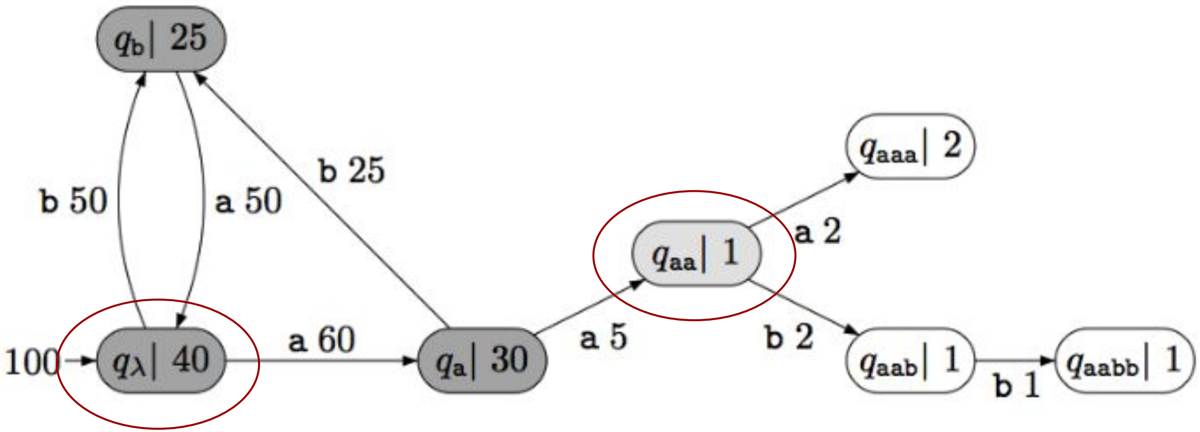
Arbre de préfixe

λ	490	abb	4	abab	2	aaaaa	1
a	128	baa	9	abba	2	aaaab	1
b	170	bab	4	abbb	1	aaaba	1
aa	31	bba	3	baaa	2	aabaa	1
ab	42	bbb	6	baab	2	aabab	1
ba	38	aaaa	2	baba	1	aabba	1
bb	14	aaab	2	babb	1	abbaa	1
aaa	8	aaba	3	bbaa	1	abbab	1
aab	10	aabb	2	bbab	1		
aba	10	abaa	2	bbba	1		



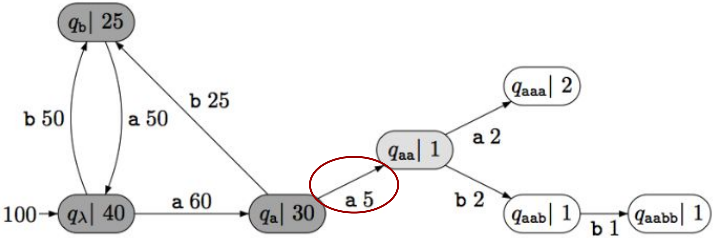
Merge & Fold

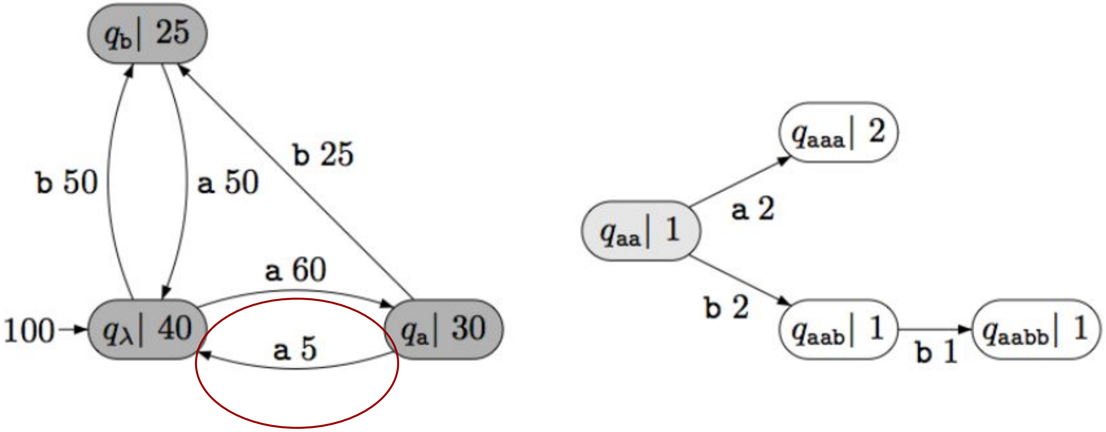
Merge & Fold entre λ et aa



Merge & Fold

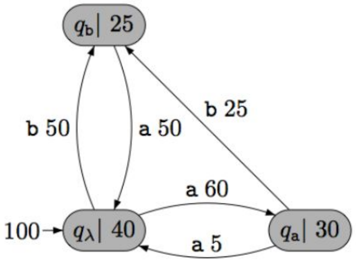
1. Déconnecter et reconnecter

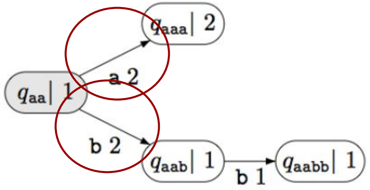


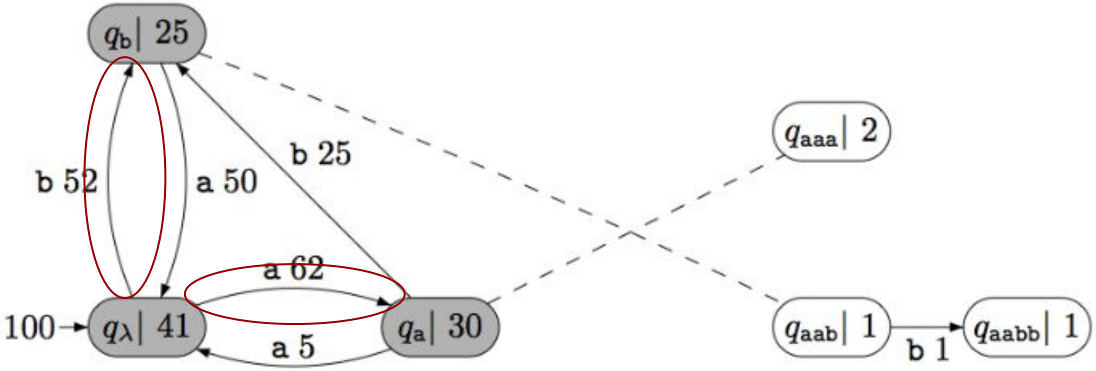


Merge & Fold

1. Déconnecter et reconnecter



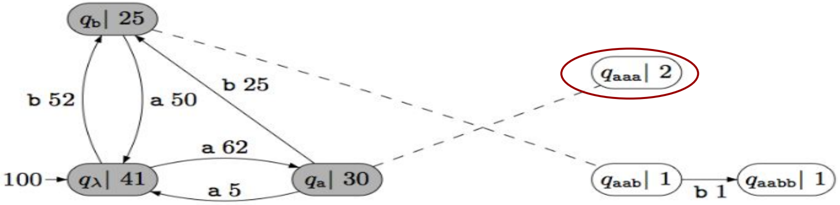


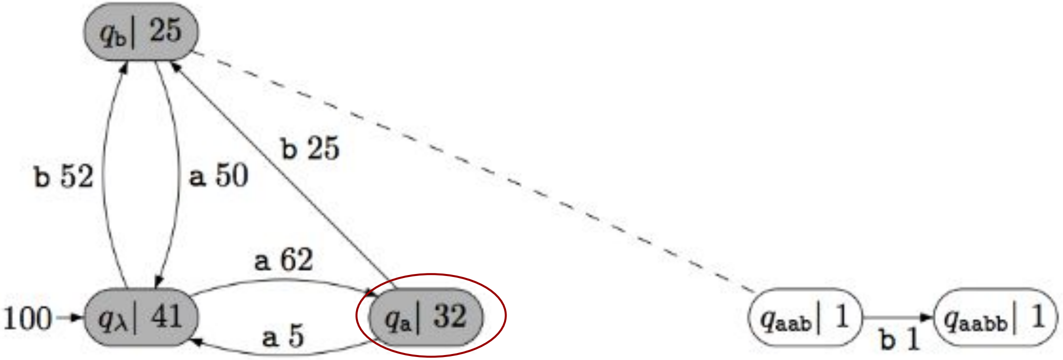


Exemple du mot aaa : aa et λ ont mergés, donc lire aaa revient à prendre la transition a à partir de λ

Merge & Fold

2. Plier

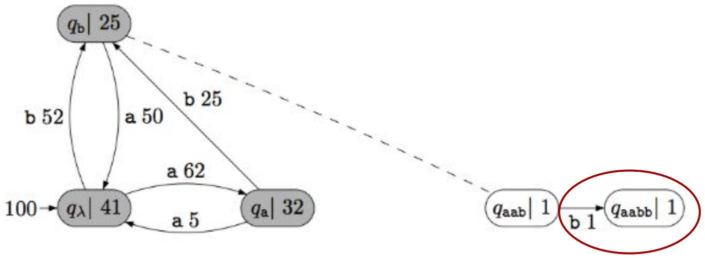


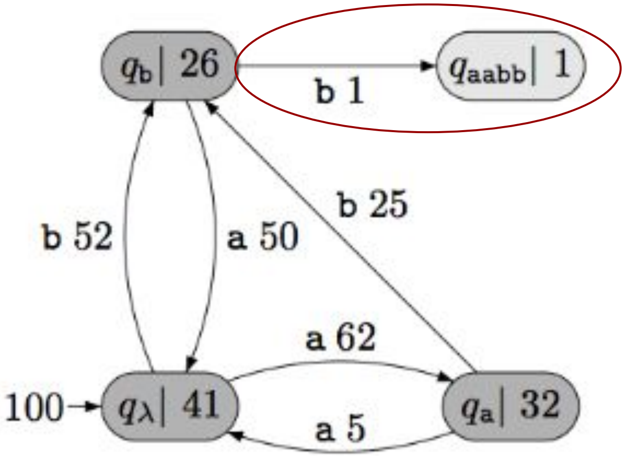


Exemple du mot aaa : tous les mots contenant aaa s'arrête après aaa. Donc l'état a à deux éléments qui s'arrêtent en plus

Inférence

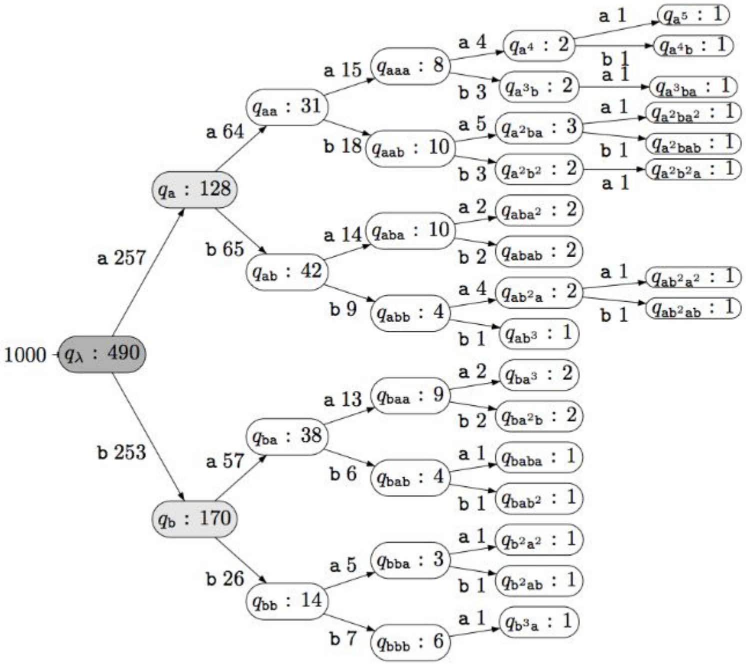
3. Coller



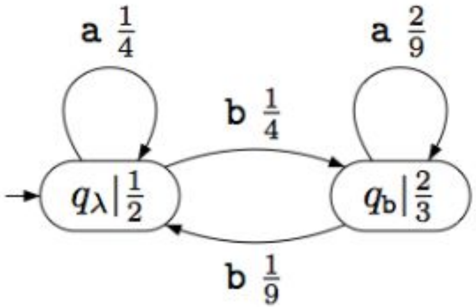


Exemple du mot aabb : après avoir merger aab, on se retrouve en b et il faut faire une transition b. Cette dernière n'existe pas donc on attache l'arbre non reconnu sur b.

Inférence



Attention à ne pas trop réduire !!!



Comment savoir quand et où utiliser Merge & Fold ?

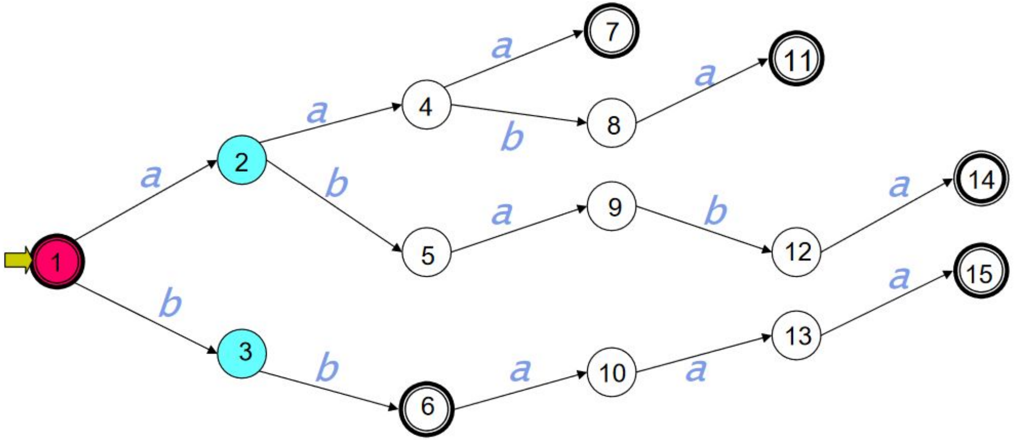
4 Mot interdit

Comment faire de l'inférence quand certains mots sont interdits ?

MDPI

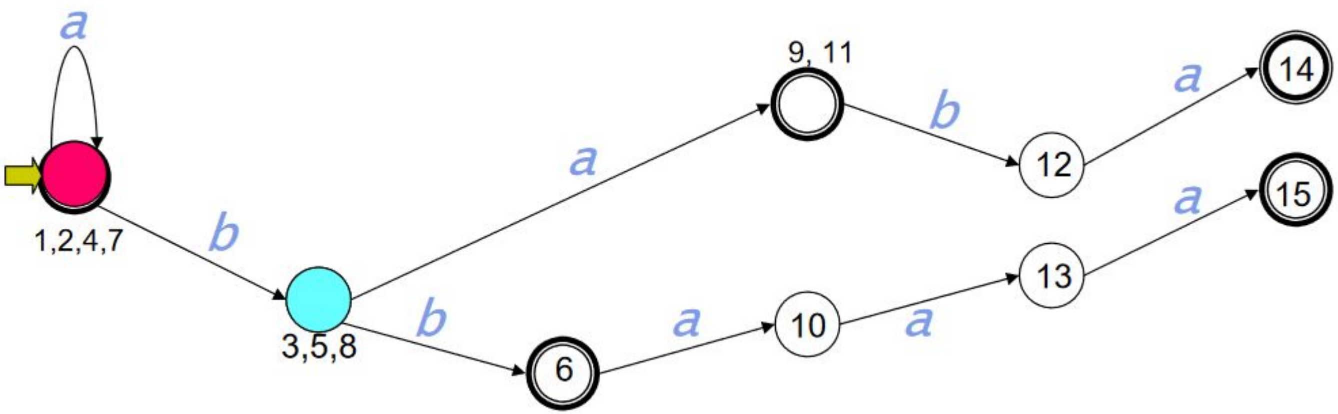
$S_+=\{\lambda, aaa, aaba, ababa, bb, bbaaa\}$

$S_-=\{aa, ab, aaaa, ba\}$



En rouge les états Validés; En bleu les états en Attente (enfants)

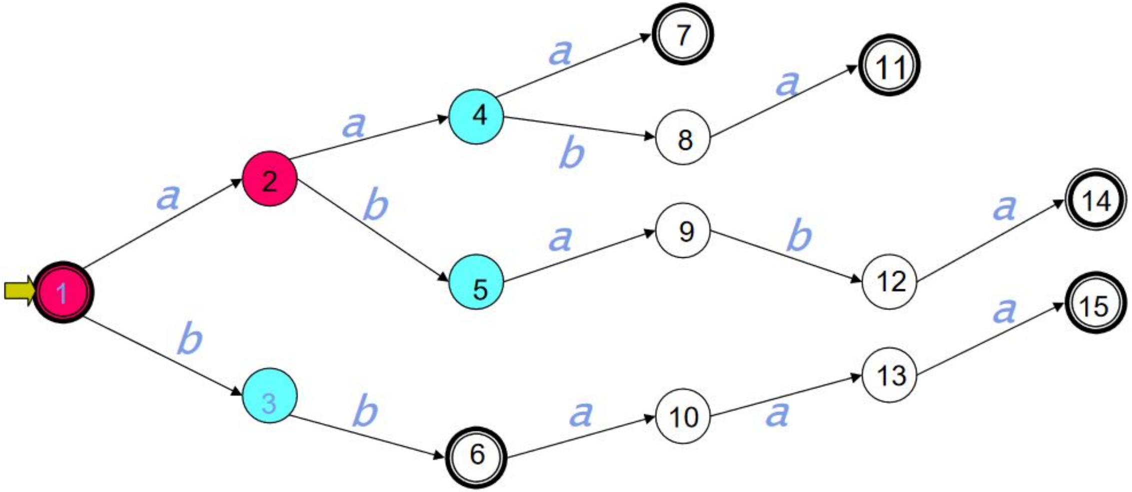
MDPI



S_{-}={ aa, ab, aaaa, ba}

Merge & Fold entre les états 1 et 2

MDPI

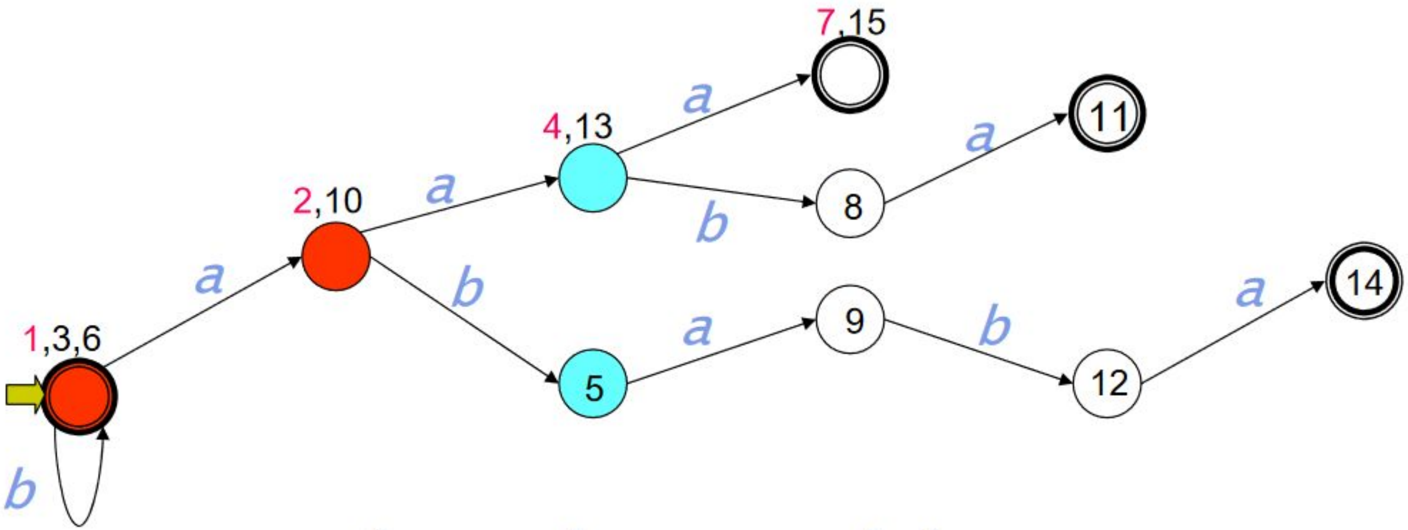


$S_- = \{aa,\ ab,\ aaaa,\ ba\}$

Aaaa est incompatible, donc 2 devient rouge

On merge 1 et 3

MDPI

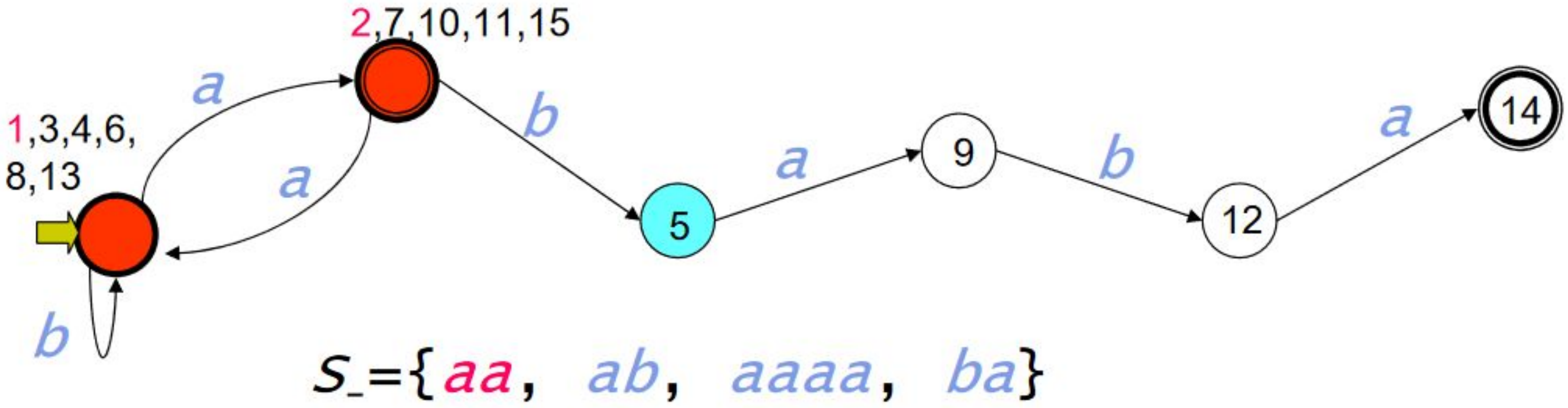


$S_-=\{aa,\ ab,\ aaaa,\ ba\}$

On merge 1 et 3 : opération réussi car pas de mots incompatibles

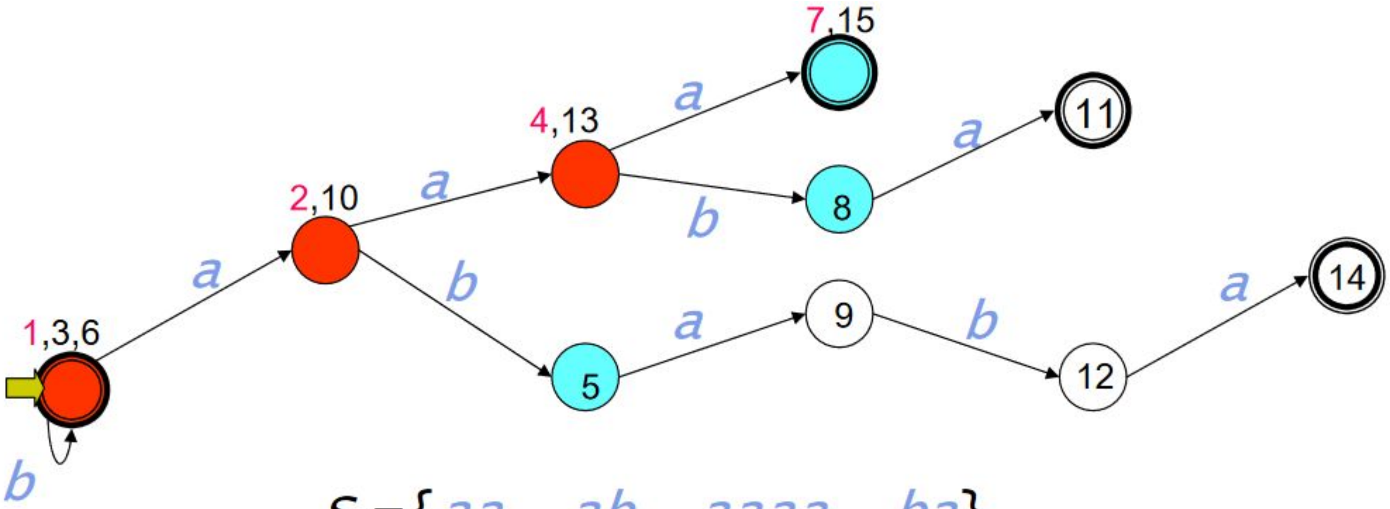
On merge {1,3,6} et {4,13}

MDPI



On merge {1,3,6} et {4,13} : mot incompatible, on test alors {2, 10} et {4,13}

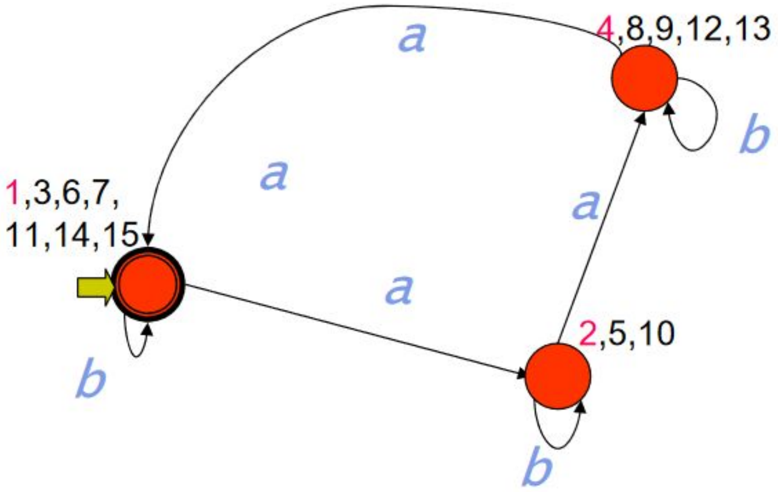
MDPI



$S_- = \{aa, ab, aaaa, ba\}$

On merge {1,3,6}, {2, 10} et {4,13} : encore un mot incompatible, {4, 13} ne peut plus être merger avec un Valide, il devient alors Validé.

MDPI



$S_{-}=\{aa, ab, aaaa, ba\}$

Une fois qu'il ne reste que des états Validés, on arrête.

3

Inférence

Comment construire un langage probabiliste à partir d'échantillons.

Inférence

Hoeffding bounds (test de compatibilité)

Soit f la fréquence d'un symbole pour un état 1 et n le nombre d' élément dans cet état (somme sortants + absorbés)

$$\gamma \leftarrow \left| \frac{f_1}{n_1} - \frac{f_2}{n_2} \right|$$

$$\gamma < \left(\sqrt{\frac{1}{n_1}} + \sqrt{\frac{1}{n_2}} \right) . \sqrt{\frac{1}{2} \ln \frac{2}{\alpha}}$$

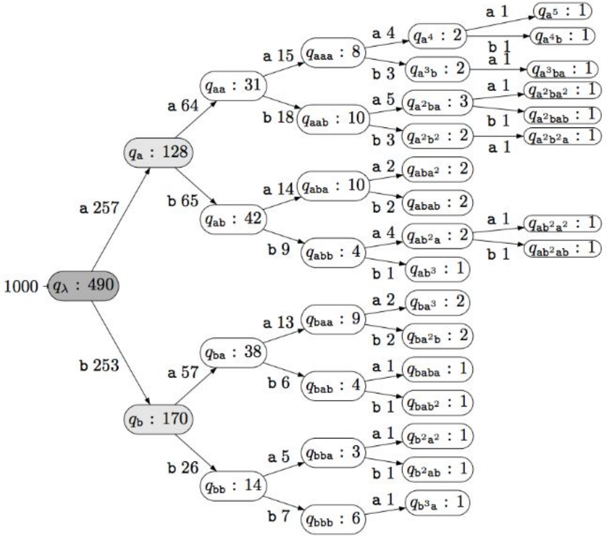
Avec α entre 0 et 1

Choisir une Limite de fréquence t_0

Inférence

Hoeffding bounds (test de compatibilité)

Peut-on merger deux états ? Tester la compatibilité sur tous les suffixes.

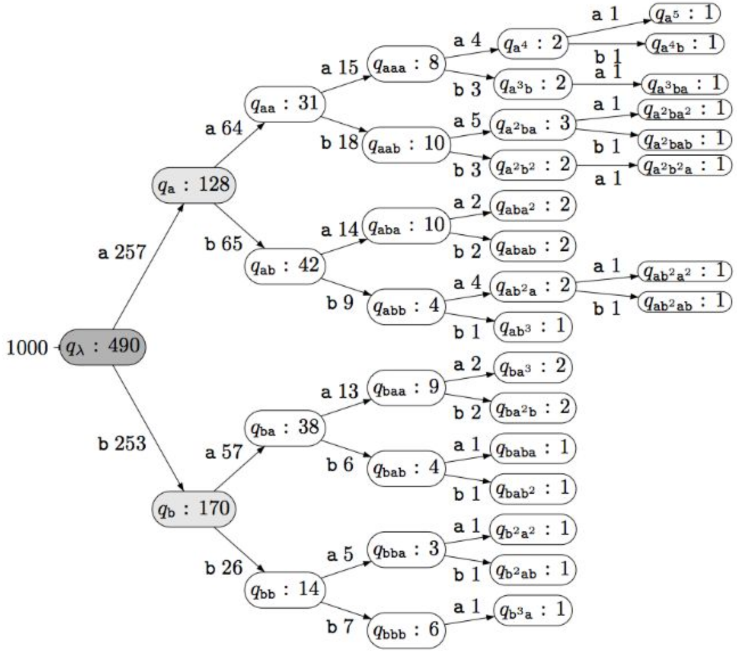


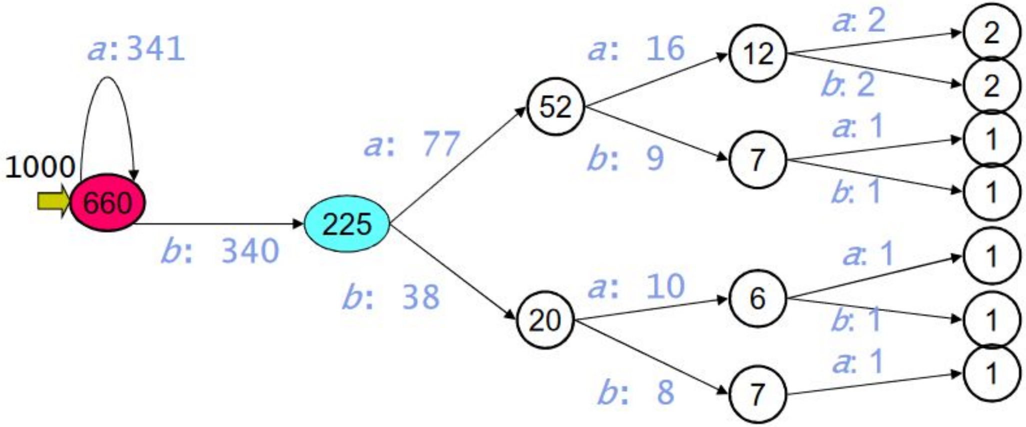
Merger λ et a = Tester Hoeffding bounds sur :
aΣ* et aaΣ* (λ et a avec le suffixe a et)
bΣ* et abΣ* (λ et a avec le suffixe b)
Les fréquences d'absorption
Et de même pour TOUS les successeurs !

257/1000 et 64/257
253/1000 et 65/257
490/1000 et 128/257 etc.
Les tests donnent Vrai avec α=0.05 et t_0=30

Inférence

ALERGIA : Hoeffding + Merge & Fold



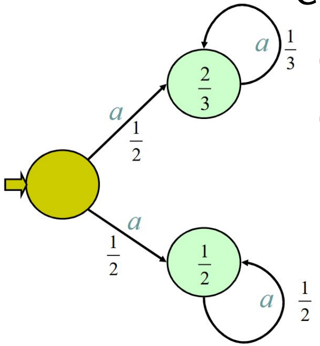


Conclusion

Monsieur, c’est pas bientôt fini tout ce charabia !?

Recherche Opérationnelle

L'inférence n'est pas qu'une question de convergence, deux critères de vraisemblance sont aussi à prendre en compte :



- Langage PAC-learnable
- Divergence de Kullback-Liebler ("inclut" dans ALERGIA) : avec A des automates

$$D(A_0||A_1) = \sum_{q_i \in Q_0} \sum_{a \in \Sigma \cup \{\#\}} c_i \; \gamma_0(q_i, a) \log \frac{\gamma_0(q_i, a)}{\gamma_1(q_{\underline{i}}, a)}$$

$$= -H(A_0) - \sum_{q_i \in Q_0} \sum_{a \in \Sigma \cup \{\#\}} c_i \; \gamma_0(q_i, a) \log \gamma_1(q_{\underline{i}}, a)$$

- Perplexité ($2^{\text{log-likelihood}}$) avec S un mot

$$LL = \left(-\frac{1}{||S||} \sum_{j=1}^{|S|} \sum_{i=1}^{|x|} \log P(x_i^j | q^i) \right)$$

Recherche Opérationnelle

Parfois les automates sont trop grand pour faire de l’inférence par ces méthodes, il faut alors passer par de la Recherche Opérationnelle :

- Algorithme génétique : les gènes sont des partitions d’états encodés en fonction du set; les opérations se font sur le numéro de set.

{{1,3} {2} {4,5}}

{{1} {2} {3,4,5}}

{{1} {2} {3} {4,5}}

{{1,2,6}{3,7,9,10}{4,8,12}{5}{11}}

(112341232253)

Recherche Opérationnelle

Parfois les automates sont trop grand pour faire de l'inférence par ces méthodes, il faut alors passer par de la Recherche Opérationnelle :

- Recherche Tabou : tester les convergences et leur attribuer une note en fonction de Kullback-Liebler. Effectuer le meilleur regroupement, annuler+Tabou si mot incompatible ou non-consistance (des zones sans états finaux).