Apprentissage séquentiel et non supervisé en entreprise

L'IA, un outil de compétitivité pour les entreprises

Formation DIRECCTE

Aurélien Garivier

Outline

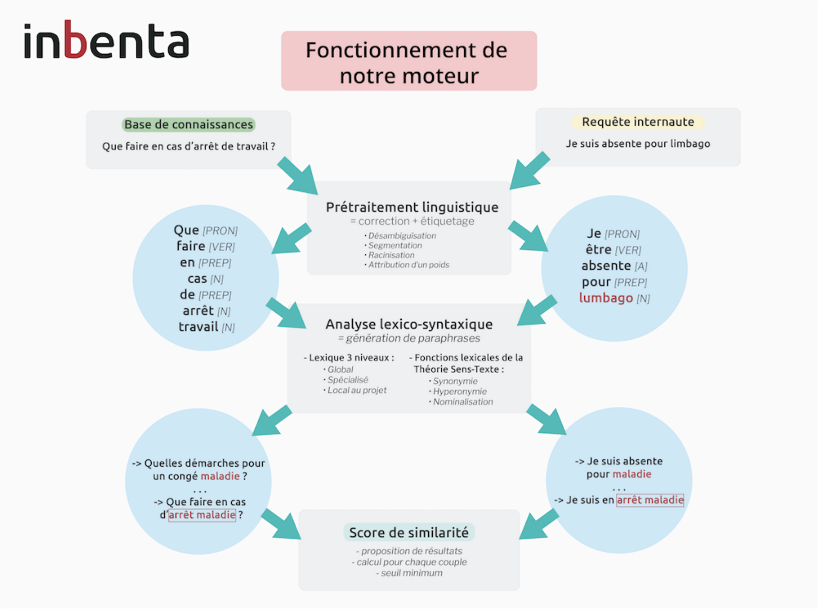
Création automatique de FAQ

Enchères automatiques

Analyse séquentielle et problèmes de bandits

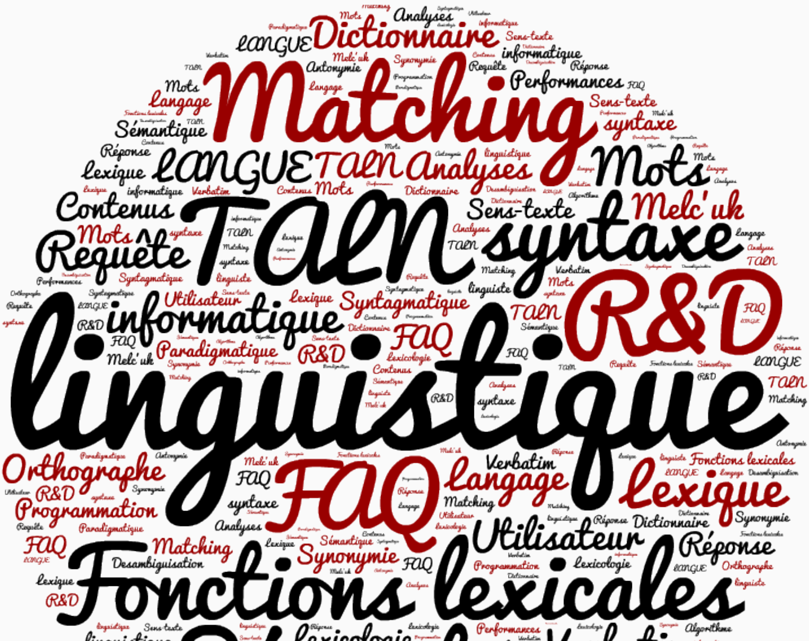
- Stratégie statistique classique
- Stratégie séquentielle
- Stratégie pleinement séquentielle: algorithmes de bandits

Inbenta

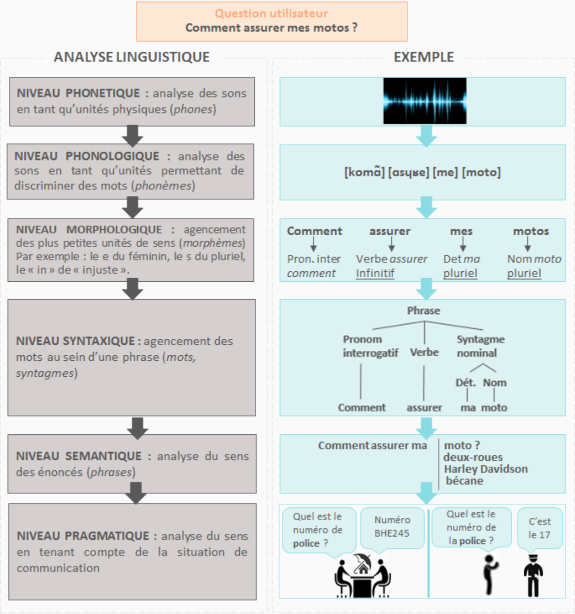


Src: Inbenta.com

Inbenta


linguistique
Matching
TALN
R&D
FAQ
Fonctions lexicales
syntaxe
Mots
LANGUE
Dictionnaire
informatique
Utilisateur
lexique
Langage
Analyses
Requête
Contenus
Orthographe
Programmation
Sémantique
Performances
Sens-texte
Melc'uk
Paradigmatique
Syntagmatique
Réponse
Synonymie
Verbatim
Lexicologie
Desambiguisation
Antonymie

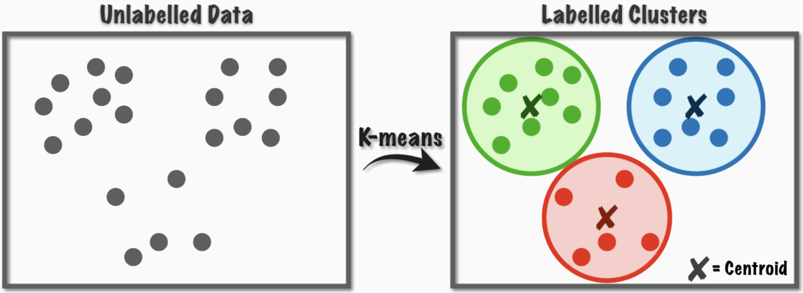
Inbenta



Src: Inbenta.com

Projet spécifique

- Compétences linguistiques fortes
- Besoin d'algorithmes personnalisés pour le contexte précis
- Création automatique d'une FAQ
- Enormément de questions, à traiter automatiquement
- Besoin d'énormément de robustesse (beaucoup de "bruit")



Src: `towardsdatascience.com`

La solution

- Les connaissances linguistiques donnent une excellente notion d'affinité entre mots
- On l'étend intelligemment à une bonne notion de distances entre questions
- Puis on met en œuvre une stratégie d'apprentissage non-supervisé basée sur les affinités
- k-means est identifiée comme la plus prometteuse, mais
  - trop chère en coût de calcul
  - trop sensible au bruit
- Une variante est développée qui est paramétrable pour les besoins spécifiques

Outline

Création automatique de FAQ

Enchères automatiques

Analyse séquentielle et problèmes de bandits

- Stratégie statistique classique
- Stratégie séquentielle
- Stratégie pleinement séquentielle: algorithmes de bandits

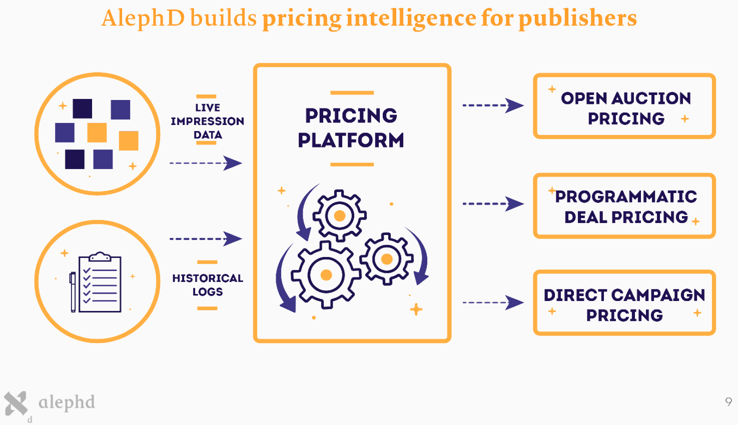
Un problème aux deux faces

- Acheteur: optimiser les bids
- Vendeur: optimiser les prix de réserve

alephd

We help publishers

Un problème aux deux faces



Outline

Création automatique de FAQ

Enchères automatiques

Analyse séquentielle et problèmes de bandits

- Stratégie statistique classique
- Stratégie séquentielle
- Stratégie pleinement séquentielle: algorithmes de bandits

Analyse séquentielle

Définition WikipediA

En statistique, l'analyse séquentielle ou le test d'hypothèse séquentiel est une analyse statistique où la taille de l'échantillon n'est pas fixée à l'avance. Plutôt, les données sont évaluées au fur et à mesure qu'elles sont recueillies, et l'échantillonnage est arrêté selon une règle d'arrêt prédéfinie, dès que des résultats significatifs sont observés. Ainsi, une conclusion peut parfois être atteinte à un stade beaucoup plus précoce que ce qui serait possible avec des tests d'hypothèse ou des estimations plus classiques, à un coût financier ou humain par conséquent inférieur.

A/B testing

Src: http://cdn1.tnwcdn.com/

A/B testing

Définition
En marketing et en Business Intelligence, l'A/B testing (ou test A/B) est la comparaison de deux versions d'une page web afin de déterminer la plus performante. Les deux versions appelées A et B sont présentées à des utilisateurs similaires, et celle qui obtient le meilleur taux de conversion est conservée.

Exemple: campagne Obama 2008 (source: WikipediA)
Quatre boutons et six médias (trois images et trois vidéos) ont été combinés de façon à obtenir 24 combinaisons différentes afin de déterminer laquelle permettait d'obtenir le taux de souscription le plus élevé.

$\Longrightarrow$ La combinaison gagnante a obtenu un taux de souscription de 11,6% alors que la page originale avait un taux de souscription de 8,26%.

“Modèle de bandit”

- Nombre total d'interactions: T
- Le système choisit de présenter au visiteur t le choix $I_t \in \{A, B\}$
  - si $I_t = A$, le feedback est $X_{A,t}$
  - si $I_t = B$, le feedback est $X_{B,t}$

  où

  $$\forall t \geq 1, \quad (X_{A,t}, X_{B,t}) \overset{iid}{\sim} \Big(\mathcal{N}(\mu_A, \sigma^2),\ \mathcal{N}(\mu_B, \sigma^2)\Big)$$

  ou n'importe quelle autre loi (par exemple Bernoulli ou Poisson) paramétrée par $\mu = (\mu_A, \mu_B)$
- **But:** Maximiser $S_T(\mu) = \sum_{t=1}^{T} X_{I_t,t}$ en espérance

Mesurer l'efficacité d'une stratégie : regret

But équivalent: minimiser le regret = ce que l'on perd par rapport à un système utilisant toujours la meilleure option

$$\begin{aligned} R_\mu(T) &= T \max\{\mu_A, \mu_B\} - \mathbb{E}_\mu\left[\sum_{t=1}^{T} X_{I_t,t}\right] \\ &= \left|\mu_A - \mu_B\right| \mathbb{E}_\mu\left[N_m(T)\right] \end{aligned}$$

où $N_X(T) = \sum_{t \leq T} \mathbb{1}\{I_t = X\}$ est le nombre de fois que l'option $X \in \{A, B\}$ a été présentée, et $m = \operatorname{argmin}_X \mu_X$

Comment faire ? approche statistique classique

Étape 1: Expérimenter

- taille d'échantillon n
- partition I_A, I_B de $\{1, \ldots, 2n\}$ telle que $|I_A| = |I_B| = n$

$\Longrightarrow$ Le visiteur k reçoit la version A si $k \in I_A$ et B sinon

- on enregistre les conversions des n visiteurs

Étape 2: Décider

- La version $X \in \{A, B\}$ ayant le meilleur taux de conversion moyen est conservée

Étape 3: Appliquer

- La version X est appliquée jusqu'à la fin

Choix de la taille n de l'échantillon
Cas 1: écart $\Delta = |\mu_A - \mu_B|$ connu

input: T, Δ

$n := \left\lceil \frac{2W\left(\frac{T^2\Delta^4}{32\pi}\right)}{\Delta^2} \right\rceil$

for $k \in \{1, \ldots, n\}$ **do**

  choose $I_{2k-1} = A$ and $I_{2k} = B$

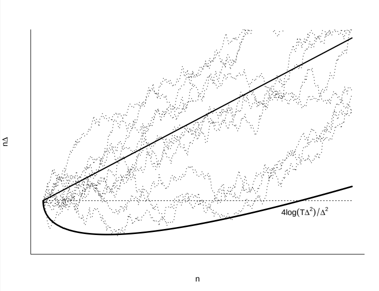
end for

$X := \mathrm{argmax}_Y \hat{\mu}_{Y,n}$

for $t \in \{2n+1, \ldots, T\}$ **do**

  choose $I_t = X$

end for



W désigne la fonction de Lambert définie pour $y > 0$ by $W(y)\exp(W(y)) = y$. Ainsi, $\overline{n} \approx 4\log(T\Delta^2)/\Delta^2$

Garantie de performance

Théorème
Pour ce choix $\overline{n}$ de taille de l'échantillon,

$$R_\mu^{\overline{n}}(T) \leq \frac{4}{\Delta} \log\left(\frac{T\Delta^2}{4.46}\right) - \frac{2}{\Delta} \log\log\left(\frac{T\Delta^2}{4\sqrt{2\pi}}\right) + \Delta$$

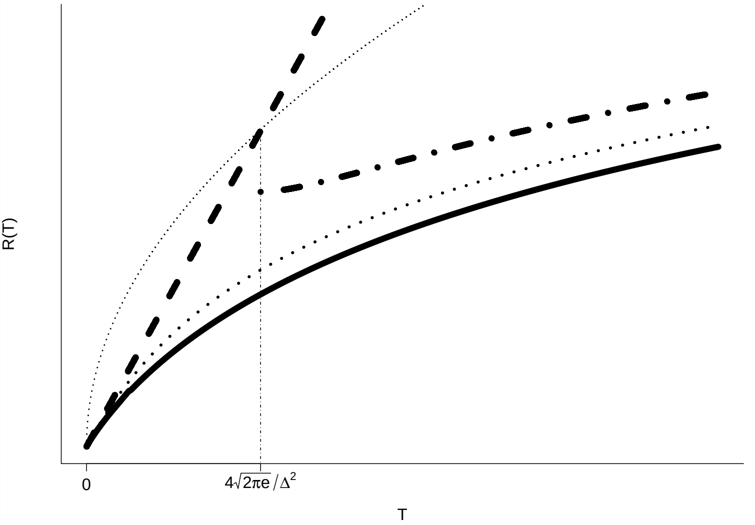
dès que $T\Delta^2 > 4\sqrt{2\pi e}$, tandis que $R_\mu^{\overline{n}}(T) \leq T\Delta/2 + \Delta$ sinon.
Dans tous les cas, $R_\mu^{\overline{n}}(T) \leq 2.04\sqrt{T} + \Delta$.

Cela est "optimal" : quand $T \to \infty$,

$$\inf_{1 \leq n \leq T} R_\mu^n(T) \sim \frac{4\log(T)}{\Delta}$$

$$\max_{\Delta} \inf_{1 \leq n \leq T} R_\mu^n(T) - \Delta \,\dot{\sim}\, \sqrt{T}$$

Performance: illustration



Choix de la taille n de l'échantillon
Cas 2: écart $\Delta = |\mu_A - \mu_B|$ inconnu

- On est obligé de se prémunir contre le "pire" des écarts (borne minimax) qui est de l'ordre de $1/\sqrt{T}$.
- C'est le cas où on peut à peine faire la différence entre les deux versions sur l'ensemble des interactions attendues
  $\implies$ on passe une fraction non négligeable du temps à expérimenter
- On ne peut alors faire mieux qu'un regret de l'ordre de

$$R_\mu(T) \dot{\sim} \sqrt{T}$$

ou bien pire si l'écart Δ est très important !

Approche statistique classique: $+/-$

1. Expérimenter / 2. Décider / 3. Appliquer

+ simplicité de conception
+ simplicité d'application
+ maîtrise théorique ancienne

- choix de la taille n de l'échantillon ?
- nécessite de connaître le nombre d'applications
- frustrant: quand l'issue de la comparaison devient clairement prévisible, on aimerait arrêter l'expérience
- inefficace !

Stratégie séquentielle

On fusionne les étapes 1 et 2 en ne fixant pas n à l'avance:

Étape 1-2: Expérimenter tant que nécessaire

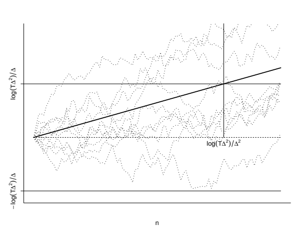
- attribution aléatoire de la version A au visiteur $2k-1$ ou $2k$
- règle d'arrêt τ: si après $2k$ visiteurs la version $X \in \{A, B\}$ apparaît significativement meilleure, on stoppe l'expérimentation

Étape 3: Appliquer

- La version X est appliquée jusqu'à la fin

Règle d'arrêt
Cas 1: écart $\Delta = |\mu_A - \mu_B|$ connu

input: T, Δ
$I_1 = A, I_2 = B$, $s := 2$
while $|\hat{\mu}_A(s) - \hat{\mu}_B(s)| < \frac{2\log\left(T\Delta^2\right)}{\Delta s}$ **do**
  choose $I_{s+1} = A$ and $I_{s+2} = B$
  $s := s + 2$
end while
$X := \mathrm{argmax}_Y \hat{\mu}_Y(s)$
for $t \in \{s+1, \ldots, T\}$ **do**
  choose $I_t = X$
end for



On stoppe quand l'écart entre les récompenses cumulées devient plus grand que $\log(T\Delta^2)/\Delta$

Garantie de performance

Théorème
Si $T\Delta^2 \geq 1$, alors la stratégie précédente vérifie

$$R_\mu(T) \leq \frac{\log(eT\Delta^2)}{\Delta} + \frac{4\sqrt{\log(T\Delta^2)} + 4}{\Delta} + \Delta\,.$$

Sinon, $R_\mu(T) \leq T\Delta/2 + \Delta$.
Dans tous les cas, $R_\mu(T) \leq 10\sqrt{T/e} + \Delta$.

Cela est "optimal" : n'importe quelle stratégie uniformément efficace sur tous les problèmes où l'écart est Δ satisfait

$$\liminf_{T\to\infty} \frac{R_\mu(T)}{\log(T)} \geq \frac{1}{\Delta} \qquad \text{et} \qquad \max_{\Delta} R_\mu(T) - \Delta \dot{\geq} \sqrt{T}\,.$$

Règle d'arrêt
Cas 2: écart $\Delta = |\mu_A - \mu_B|$ inconnu

input: T

$I_1 = A, I_2 = B$, $s := 2$

while $|\hat{\mu}_A(s) - \hat{\mu}_B(s)| < \sqrt{\frac{8\log(T/s)}{s}}$ **do**

  choose $I_{s+1} = A$ and $I_{s+2} = B$

  $s := s + 2$

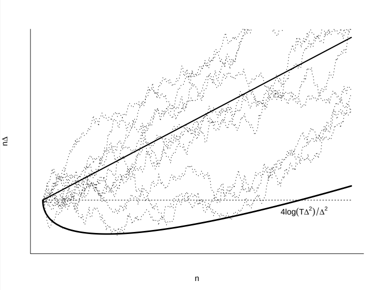
end while

$X := \mathrm{argmax}_Y \hat{\mu}_Y(s)$

for $t \in \{s+1, \ldots, T\}$ **do**

  choose $I_t = X$

end for



On stoppe quand l'écart réduit entre les deux moyennes empiriques devient significatif.

Garantie de performance

Théorème
Si $T\Delta^2 > 4e^2$, la statégie précédente satisfait:

$$R_\mu(T) \leq \frac{4\log\left(\frac{T\Delta^2}{4}\right)}{\Delta} + \frac{334\sqrt{\log\left(\frac{T\Delta^2}{4}\right)}}{\Delta} + \frac{178}{\Delta} + \Delta.$$

Sinon, $R_\mu(T) \leq T\Delta$.
Dans tous les cas, $R_\mu(T) \leq 32\sqrt{T} + \Delta$.

Cela est "optimal" : n'importe quelle stratégie uniformément efficace sur tous les problèmes où l'écart est Δ satisfait

$$\liminf_{T\to\infty} \frac{R_\mu(T)}{\log(T)} \geq \frac{4}{\Delta} \qquad \text{et} \qquad \max_{\Delta} R_\mu(T) - \Delta \dot{\geq} \sqrt{T}\ .$$

Stratégie séquentielle: +/−

1-2. Expérimenter tant que nécessaire / 3. Appliquer

+ plus satisfaisant pour l'intuition (expérience arrêtée dès que possible)
+ on garde la simplicité d'un choix définitif
+ théorie statistique établie

– théorie statistique moins connue
– longueur de l'expérimentation inconnue
– nécessite de connaître le nombre d'applications
– on peut encore faire mieux!

Stratégie pleinement séquentielle

On fusionne les étapes 1,2 et 3:

Étape 1-2-3: Explorer et Exploiter

- une règle d'échantillonnage indique, à chaque instant, quelle option attribuer
- cette règle doit, pour chaque visiteur k, trouver un équilibre entre exploration des deux options et exploitation des données accumulées jusqu'au visiteur $k - 1$

Les principales stratégies pleinement séquentielles

- La plus intuitive: ϵ-greedy
- Remarque: le plug-in ne marche pas du tout !
- UCB remplace l'estimateur par une borne de confiance (cf infra)
- Politique randomisée EXP3: maintient une loi de probabilité sur les options
- Politique randomisée Bayésienne: Thompson Sampling
- Remarque: elles ne nécessitent pas de connaître l'horizon

Le paradigme optimiste – “Théorie du Wishful Thinking”

Algorithmes **optimistes** : [Lai&Robins ’85; Agrawal ’95]

Fais comme si tu te trouvais dans l’environnement qui t’est le plus favorable parmi tous ceux qui rendent les observations suffisamment vraisemblables

D’abord présenté dans un contexte bandit, puis largement généralisé ces dernières années.

Propriétés

De façon plutôt inattendue, les méthodes optimistes se révèlent :

- pertinentes dans des cadres très différents
- efficaces
- robustes
- simples à mettre en oeuvre

Explication intuitive:

- soit le modèle optimiste est bon, et on agit bien;
- soit il est mauvais, et on réduit bien l'incertitude.

Exemple de stratégie purement séquentielle: UCB (Upper Confidence Bound)

Stratégie optimiste: on remplace l'estimée par une borne supérieure de confiance

---

1: **input:** T

2: **for** $t \in \{1, \ldots, T\}$ **do**

3: $\quad I_t = \mathrm{argmax}_{X \in \{A,B\}} \hat{\mu}_X(t-1) + \sqrt{\frac{2}{N_X(t-1)} \log\left(\frac{T}{N_X(t-1)}\right)}$

4: **end for**

$\Longrightarrow$ résoud le *dilemme exploration/exploitation*

Garantie de performance

Théorème
Pour tout $\epsilon \in (0, \Delta)$ tq $T(\Delta - \epsilon)^2 \geq 2$ et $T\epsilon^2 \geq e^2$, le regret de la stratégie précédente est borné par

$$R_\mu(T) \leq \frac{2\log\left(\frac{T\Delta^2}{2}\right)}{\Delta\left(1-\frac{\epsilon}{\Delta}\right)^2} + \frac{2\sqrt{\pi\log\left(\frac{T\Delta^2}{2}\right)}}{\Delta\left(1-\frac{\epsilon}{\Delta}\right)^2} + \Delta\left(\frac{30e\sqrt{\log(\epsilon^2 T)}+16e}{\epsilon^2}\right) + \frac{2}{\Delta\left(1-\frac{\epsilon}{\Delta}\right)^2} + \Delta.$$

Ainsi, $\limsup_{T\to\infty} \frac{R_\mu(T)}{\log(T)} \leq \frac{2}{\Delta}$. De plus, $R_\mu(T) \leq 33\sqrt{T} + \Delta$.

Cela est "optimal" : n'importe quelle stratégie uniformément efficace sur tous les problèmes où l'écart est Δ satisfait

$$\liminf_{T\to\infty} \frac{R_\mu(T)}{\log(T)} \geq \frac{2}{\Delta} \qquad \text{et} \qquad \max_{\Delta} R_\mu(T) - \Delta \gtrsim \sqrt{T}\ .$$

Stratégies purement séquentielles: $+/-$

1-2-3. Explorer et Exploiter

- $+$ optimal pour minimiser le regret
- $+$ ne nécessite pas de connaître l'ordre de grandeur du nombre d'applications
- $+$ très en vogue en machine learning (ex: COLT)

- – encore minoritaire dans la communauté statistique, peu diffusé
- – pas de fin d'expérimentation ni de décision claire
- – on conserve tout le temps les deux versions

Résumé: efficacité relative des stratégies

Théorème: pour des stratégies optimales dans leur catégorie,

$$R_\mu(T) \sim \frac{C \log(T)}{\left|\mu_A - \mu_B\right|}$$

avec C valant :

	Classique	Seq	Plein^t Seq
$\left\|\mu_A - \mu_B\right\|$ connu	4	1	1/2
$\left\|\mu_A - \mu_B\right\|$ inconnu	∞	4	2