Algorithme de bandit et obsolescence : un modèle pour la recommandation

Algorithme de bandit et obsolescence : un modèle pour la recommandation

Bandits classiques :
- un nombre fixé de choix disponibles (les bras) ;
- les réponses ne varient pas au cours du temps.

Algorithme de bandit et obsolescence : un modèle pour la recommandation

Bandits classiques :
- un nombre fixé de choix disponibles (les bras) ;
- les réponses ne varient pas au cours du temps.

Limitations sévères, en particulier pour les moteurs de recommandations :
- de nouveaux items apparaissent ;
- les anciens perdent de l'attractivité.

Cadre :
- de nouveaux bras apparaissent de manière régulière (flux de bras) ;
- décroissance exponentielle de l'ensemble des bras.

Cadre :
- de nouveaux bras apparaissent de manière régulière (flux de bras) ;
- décroissance exponentielle de l'ensemble des bras.

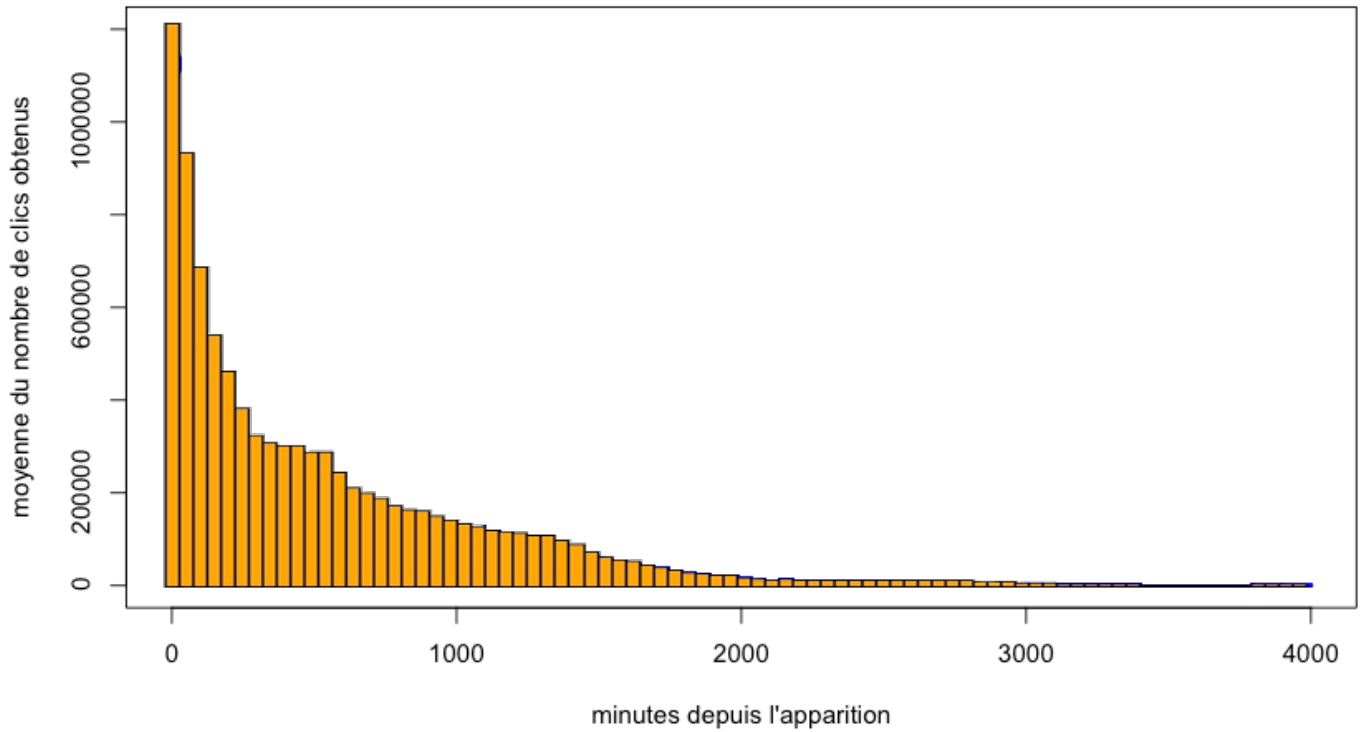

Figure – Obsolescence de la popularité des bras. Données issues des 100 documents les plus cliqués du challenge CLEF-NEWSREEL

Déroulement :
- une succession de K périodes de taille fixe L ;

Déroulement :
- une succession de K périodes de taille fixe L ;
- un nouveau bras a entre en jeu au début de chaque période,

on note t_a son instant d'apparition et p_a l'espérance de sa récompense ;

Déroulement :
- une succession de K périodes de taille fixe L ;
- un nouveau bras a entre en jeu au début de chaque période,
on note t_a son instant d'apparition et p_a l'espérance de sa récompense ;
- à chaque instant t, l'agent choisit un bras A_t parmi ceux disponibles et reçoit une récompense $Z_t \in \{0, 1\}$.

Déroulement :
- une succession de K périodes de taille fixe L ;
- un nouveau bras a entre en jeu au début de chaque période,
on note t_a son instant d'apparition et p_a l'espérance de sa récompense ;
- à chaque instant t, l'agent choisit un bras A_t parmi ceux disponibles et reçoit une récompense $Z_t \in \{0, 1\}$.

Les espérances associées aux bras décroissent à la même vitesse, selon le facteur d'obsolescence τ :

Déroulement :
- une succession de K périodes de taille fixe L ;
- un nouveau bras a entre en jeu au début de chaque période,
on note t_a son instant d'apparition et p_a l'espérance de sa récompense ;
- à chaque instant t, l'agent choisit un bras A_t parmi ceux disponibles et reçoit une récompense $Z_t \in \{0, 1\}$.

Les espérances associées aux bras décroissent à la même vitesse, selon le facteur d'obsolescence τ :
- un bras qui n'est pas optimal à l'instant t ne le sera jamais par la suite ;
- un bras optimal le reste sur toute la période.

Déroulement :
- une succession de K périodes de taille fixe L ;
- un nouveau bras a entre en jeu au début de chaque période,
on note t_a son instant d'apparition et p_a l'espérance de sa récompense ;
- à chaque instant t, l'agent choisit un bras A_t parmi ceux disponibles et reçoit une récompense $Z_t \in \{0, 1\}$.

Les espérances associées aux bras décroissent à la même vitesse, selon le facteur d'obsolescence τ :
- un bras qui n'est pas optimal à l'instant t ne le sera jamais par la suite ;
- un bras optimal le reste sur toute la période.

Lors de son apparition, nous supposons qu'un bras a une espérance supérieure à une certaine valeur η fixée a priori, cela induit qu'un bras apparu il y a plus de $\tau \log \frac{1}{\eta}$ instants ne peut pas être le bras optimal.

Déroulement :
- une succession de K périodes de taille fixe L ;
- un nouveau bras a entre en jeu au début de chaque période,
on note t_a son instant d'apparition et p_a l'espérance de sa récompense ;
- à chaque instant t, l'agent choisit un bras A_t parmi ceux disponibles et reçoit une récompense $Z_t \in \{0, 1\}$.

Les espérances associées aux bras décroissent à la même vitesse, selon le facteur d'obsolescence τ :
- un bras qui n'est pas optimal à l'instant t ne le sera jamais par la suite ;
- un bras optimal le reste sur toute la période.

Lors de son apparition, nous supposons qu'un bras a une espérance supérieure à une certaine valeur η fixée a priori, cela induit qu'un bras apparu il y a plus de $\tau \log \frac{1}{\eta}$ instants ne peut pas être le bras optimal.

L'espérance d'un bras lors de son apparition est estimée par :
$\widehat{p}_a(t) = \frac{1}{N_a(t)} \sum_{s=1}^{t} \left(Z_s \exp\left(\frac{s-t_a}{\tau}\right) \mathbb{1}_{A_s=a} \right)$
avec $N_a(t)$ le nombre de fois où le bras a a été joué aux t premiers instants.

L'espérance à l'instant t est estimée par :

$$\widehat{\mu}_a(t) = \widehat{p}_a(t) \exp\left(-\frac{t - t_a}{\tau}\right)$$

L'espérance à l'instant t est estimée par :

$$\widehat{\mu}_a(t) = \widehat{p}_a(t) \exp\left(-\frac{t - t_a}{\tau}\right)$$

A chaque instant, on joue le bras qui a le plus fort UCB :

$$U_a(t) = \widehat{\mu}_a(t) + \sqrt{\frac{2 \log\left(\tau \log \frac{1}{\eta}\right)}{N_a(t)}}$$

L'espérance à l'instant t est estimée par :

$$\widehat{\mu}_a(t) = \widehat{p}_a(t) \exp\left(-\frac{t - t_a}{\tau}\right)$$

A chaque instant, on joue le bras qui a le plus fort UCB :

$$U_a(t) = \widehat{\mu}_a(t) + \sqrt{\frac{2 \log\left(\tau \log \frac{1}{\eta}\right)}{N_a(t)}}$$

Le regret cumulé moyen $R(T)$ de l'algorithme F-UCB vérifie :

$$R(T) \leq 8 \log\left(\tau \log \frac{1}{\eta}\right) \sum_a \frac{1}{\Delta_a^2} + (K-1)\left(1 + \frac{2}{\tau \log \frac{1}{\eta}}\right)$$

en sommant sur les bras a sous-optimaux, Δ_a est le gap minimal entre les espérances du bras optimal et du bras a.

L'hypothèse est qu'au début d'une période, les bras déjà introduits ont été joués un nombre de fois suffisant pour avoir des estimations de leurs popularités assez précises.

L'hypothèse est qu'au début d'une période, les bras déjà introduits ont été joués un nombre de fois suffisant pour avoir des estimations de leurs popularités assez précises.

Au début du round r, le bras a_r est introduit et la stratégie consiste à jouer ce bras tant qu'il est impossible de certifier que son espérance est plus faible.

L'hypothèse est qu'au début d'une période, les bras déjà introduits ont été joués un nombre de fois suffisant pour avoir des estimations de leurs popularités assez précises.

Au début du round r, le bras a_r est introduit et la stratégie consiste à jouer ce bras tant qu'il est impossible de certifier que son espérance est plus faible.

A chaque instant t de ce round, on effectue l'action :

Si $\max_{a \in A} U_a(t) > U_{a_r}(t)$ alors

Jouer $A_t = \text{argmax}_{a \in A} \widehat{\mu}_a(t)$

Sinon

Jouer $A_t = a_r$

Fin

L'hypothèse est qu'au début d'une période, les bras déjà introduits ont été joués un nombre de fois suffisant pour avoir des estimations de leurs popularités assez précises.

Au début du round r, le bras a_r est introduit et la stratégie consiste à jouer ce bras tant qu'il est impossible de certifier que son espérance est plus faible.

A chaque instant t de ce round, on effectue l'action :

Si $\max_{a \in A} U_a(t) > U_{a_r}(t)$ alors

Jouer $A_t = \text{argmax}_{a \in A} \widehat{\mu}_a(t)$

Sinon

Jouer $A_t = a_r$

Fin

La principale différence réside dans le choix du bras joué une fois le bras a_r éliminé :

- *TA-UCB* joue le bras avec l'espérance la plus forte ;
- *F-UCB* joue le bras avec la borne de confiance supérieure la plus forte.

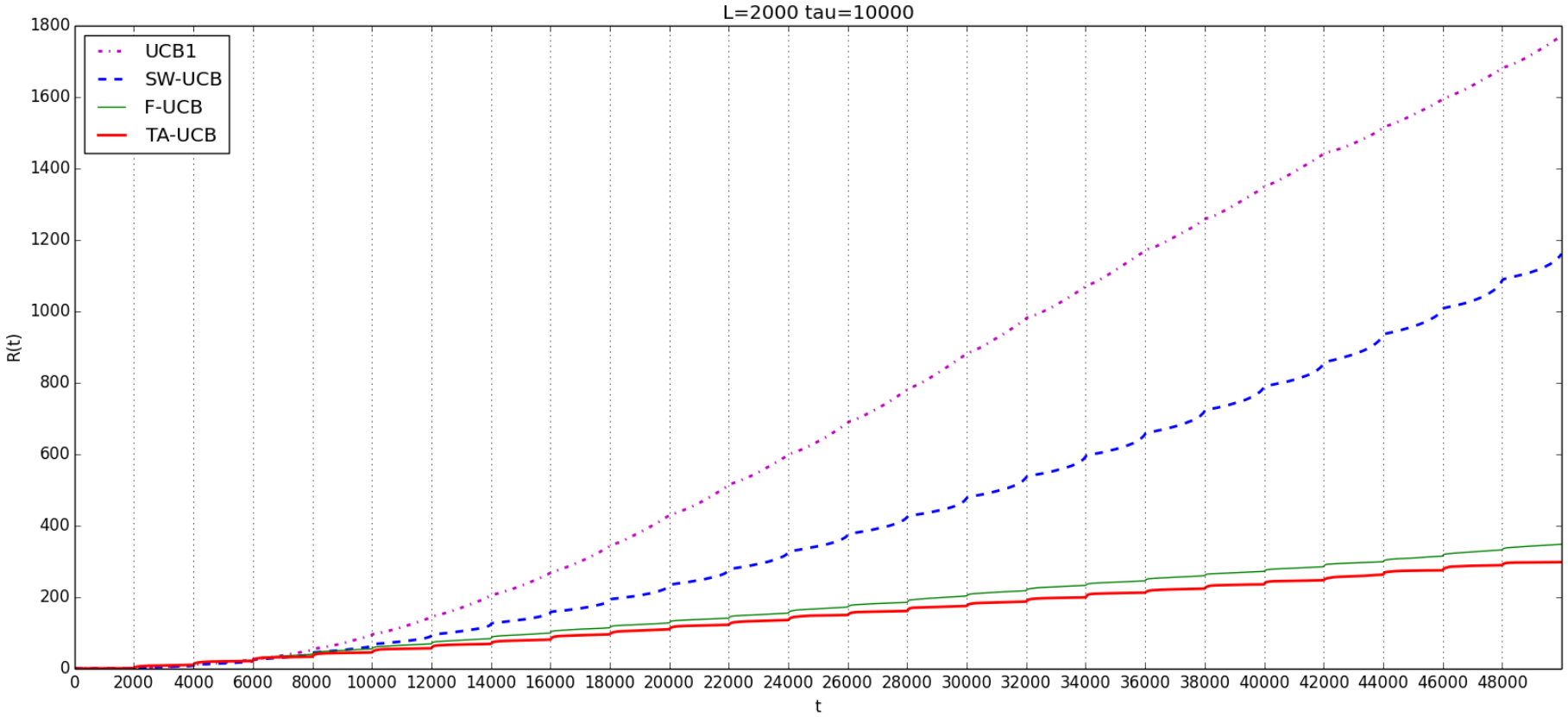

Figure – Simulation avec $L = 2000$ et $\tau = 5L$