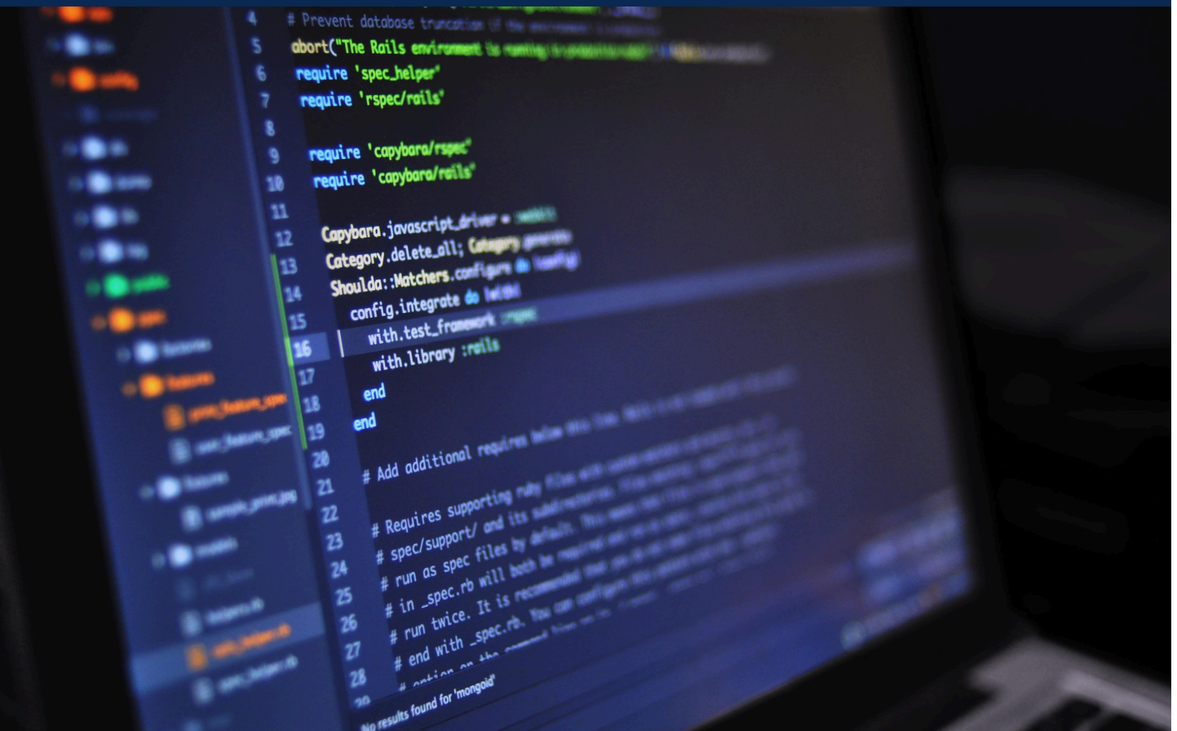
EXPLORING SCENARIOS OF UNCERTAINTY ABOUT THE USERS' PREFERENCES IN INTERACTIVE RECOMMENDATION SYSTEMS

VICENTE LYON
BORJA ERRÁZURIZ
ANTONIO LARRAIN

INDICE DE CONTENIDOS



1/MOTIVACIÓN Y CONTEXTO

CONTEXTO

- Los SRI juegan un papel fundamental en la experiencia del usuario en plataformas modernas
- Funcionan con Contextual Bandits: items como brazos para jalar y feedback como recompensa
- Tradeoff explotar mejor brazo o explorar uno nuevo

PROBLEMA Y MOTIVACIÓN

- La personalización de los SRI depende directamente de la información disponible de ese usuario
- 2 principales escenarios de incertidumbre:
  - **Cold start** para usuarios nuevos
  - **Recomendaciones erróneas continuas** por previas suposiciones incorrectas

PROPUESTA

- La propuesta es ocupar **Active Learning** (AL) para atacar estos problemas
- **AL**: identificar data points más informativos y obtener feedback de ellos para mejorar el modelo
- Ocupando una mezcla de **entropía** y **popularidad** se pude reducir el impacto de ambos escenarios.

2/ TRABAJOS RELACIONADOS

1/ MAB

2/ CONTEXTUAL BANDITS

3/ UNCERTAINTY APROACHES

ESTADO DEL ARTE

- Multi Armed Bandits: representar decisiones como 3-tupla ⟨A, Q, R⟩
- El agente busca elegir una acción (A), basada en su función de valor (Q) para maximizar la recompensa (R) en cada trial $t \in T$
- Se busca maximizar la recompensa total después de T trials
- Cada elección depende de la policy π, teniendo en cuenta que $Q_t(a) = E[r_t | a]$:
  - Acción con más valor (explotar)
  - Buscar nueva opción (explorar)

1. ϵ-Greedy: enfoque en corto plazo
2. UCB: exploración random
3. TS: Explorar acciones con mayor incertidumbre

- **Limitaciones:** no son intrínsecamente personalizados para cada usuario y, cuando hay una gran cantidad de datos y opciones, pueden tardar mucho en identificar las mejores alternativas.

1/ MAB

2/ CONTEXTUAL BANDITS

3/ UNCERTAINTY APROACHES

ESTADO DEL ARTE 

- Algoritmo incorpora las preferencias del usuario en cada interacción.
- Información explícita de usuarios: demográfica, geográfica, etc
- Información explícita de items: descripción, categorías, etc
- *Probilistic Matrix Factorization* modela usuarios e items con factores no observables

1. *LinUCB* (user and item info)
2. *FactorUCB* (user and item info, PMF)
3. *hLinUCB* (user and item info, PMF)
4. *CoLin* (user and item info, PMF)
5. *Linear UCB* (PMF)
6. *GLM-UCB* (PMF)

- **Limitaciones:** Cuando la información del usuario es inexistente o insuficiente, pueden ser ineficientes.

1/ MAB

2/ CONTEXTUAL BANDITS

3/ UNCERTAINTY APROACHES

ESTADO DEL ARTE 🎨

- Mitigar la incertidumbre es uno de los desafíos más grandes
- Métodos de Interferencia Bayesiana, los cuales son variaciones del *Thompson Sampling* aplicando una distribución *Beta.* De esta forma se ajusta la probabilidad de éxito/fallo en cada brazo
- Otro aproach es el Meta Learning, modelos que se adaptan con pocos ejemplos.

1. PTS
2. Cluster Bandit
3. ICTRTS
4. NCIF

3/ SOLUCIÓN PROPUESTA

COLD START

- Problema clásico en los sistemas recomendadores: **no se tiene información de un usario o item nuevo**
- No se pueden crear los vectores contextuales pu (preferencias de usuario) y qi (features de los items) que se utilizan para generar las recomendaciones
- Se asume la preferencia del usuario con una **constante** $c \geqslant 0$

$$\textsc{Prediction Rule:}\ \ i^*(t) = \arg\max_{i \in I}\ p_u^\top q_i$$

$$(t=0)\begin{cases} p_u = \{0 : z \in Z\} \rightarrow i^*(0) = \arg\max_{i \in I}\ \mathbf{0} \cdot q_i & \text{(random)} \\ p_u = \{c : z \in Z\} \rightarrow i^*(0) = \arg\max_{i \in I}\ \boldsymbol{c} \cdot q_i & \text{(biased)} \end{cases}$$

- El primer caso representa una **exploración** pura y el segundo una **explotación** pura

MISSLEADING ASSUMPTIONS

- Ocurre cuando el sistema aprende preferencias erróneas, lo que produce que realice recomendaciones poco precisas
- Esto se puede deber a 3 principales razones:
  - Un escenario de cold start, con pocas interacciones del usuario
  - Usuarios con preferencias muy cambiantes
  - Cuando se tienen cuentas compartidas por más de 1 persona
- Todo esto produce que el modelo haga malas suposiciones, aprenda mal y termine realizando recomendaciones malas

ACTIVE LEARNING

- Busca mejorar el proceso de entrenamiento en el Machine Learning, seleccionando los data points más informativos para el aprendizaje.
- En RecSys, se usa con el fin de seleccionar items que tengan **mayor potencial de ser calificados** por los usuarios
- Este paper se enfoca en una estrategia **no personalizada de AL**, para poder funcionar en escenarios de alta incertidumbre. Para esto utiliza dos conceptos:
  - **Popularidad**: recomendar items populares para maximizar la probabilidad de ser calificado (explotación)
  - **Entropía**: busca aumentar la cantidad de información a obtener si es que el usuario califica el item (exploración)

ACTIVE LEARNING

- **Popularidad**: se define como ϱ y es medida como la cantidad de usuarios distintos que calificaron un item
- **Entropía**: se define como ϕ y es calculada con la siguiente ecuación:

$$\phi(i) = \sum_{r} -P(r|i) \cdot \log P(r|i)$$

- En donde $P(r|i)$ es la probabilidad de recibir un rating r (1-5) dado un item i

PROPUESTA:
ESCENARIO 1

- Utilizar información disponible de los items para hacer un warm start en las primeras interacciones del usuario (t=0)
- Esto funciona para computar los features iniciales, después el modelo actualizará los valores según corresponda

$$i^*_{(t=0)} = \arg\max_{i \in I} p_u^\top q_i \quad \equiv \quad i^*_{(t=0)} = \arg\max_{i \in I} \log \rho_i \cdot \phi_i$$

PROPUESTA:
ESCENARIO 1

- Aun en un escenario de cold start, el vector de features de los ítems $qi \in Q$ se puede medir, pero lo que falta es estimar el vector pu
- Se define función de pérdida para encontrarlo:

$$f(x) = \sum_{i \in I} (y_i - x^T \cdot q_i)^2, \quad \text{where: } y_i = \log \rho_i \cdot \phi_i$$

- Se aplica el método cuasi-Newon BFGS para estimar el pu después de n iteraciones
- En cada nueva iteración se busca minimizar la función de pérdida mediante el método del gradiente

$$\text{previous one: } x_{n+1} = x_n - [H(x_n)]^{-1} \nabla f(x_n):$$

$$n \text{ iterations} \begin{cases} \overrightarrow{x_1} \leftarrow \overrightarrow{x_0} - [H(\overrightarrow{x_0})]^{-1} \nabla f(\overrightarrow{x_0}) \\ \overrightarrow{x_2} \leftarrow \overrightarrow{x_1} - [H(\overrightarrow{x_1})]^{-1} \nabla f(\overrightarrow{x_1}) \\ \vdots \\ x \leftarrow \overrightarrow{x_{n-1}} - [H(\overrightarrow{x_{n-1}})]^{-1} \nabla f(\overrightarrow{x_{n-1}}) \end{cases}$$

PROPUESTA:
ESCENARIO 2

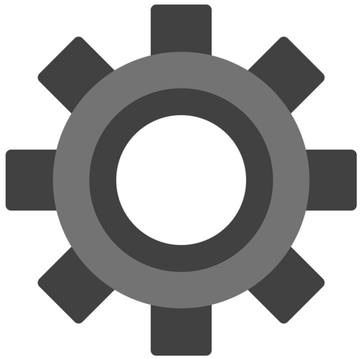
- Se define un umbral T > 0, y cuando la cantidad de recomendaciones malas consecutivas (m) sea mayor al umbral, se utiliza el aproach de AL, sino, el modelo funciona exactamente igual
- Cuando el modelo no ha hecho buenas recomendaciones durante un tiempo, el estudio asume que el usuario deja de esperar recomendaciones personalizadas
- Se debe buscar el T óptimo mediante experimentación

$$\textsc{Prediction Rule}$$

$$i^*(t) = \begin{cases} \text{original } Q_t, & \text{if } m < \mathcal{T} \\ \arg\max_{i \in I} \log \rho_i \cdot \phi_i, & \text{otherwise} \end{cases}$$

PROPUESTA: ALGORTIMOS MODIFICADOS

- 3 algoritmos modificados, uno de cada tipo de contextual bandits:
  - Linear ϵ-Greedy:
  - LinUCB
  - PTS



Algorithm 1 CONTEXTUAL ϵ-GREEDY

Require: features $Q = \{q_1, ..., q_n\}$ from PMF, popularity ρ, entropy ϕ, variance λ_p, ϵ, and the threshold $\mathcal{T} > 0$

$\mathbf{X} \leftarrow BFGS_{\mathbf{x}} \sum_i (\log \rho_i \cdot \phi_i - \mathbf{x}^\top Q_i)^2$
$\Sigma_{u,t} \leftarrow \lambda_p I_d$
$m \leftarrow 0$
for $t = 1, 2, ..., T$ **do**
  Estimates $p_{u,t} \leftarrow \Sigma_{u,t}^{-1} \cdot \mathbf{X}$
  With probability 1 - ϵ
    If $m < \mathcal{T}$:
      $i_t{}^* \leftarrow \arg\max_{i \in I \setminus R} p_{u,t}^\top q_i$
    Otherwise:
      $i_t{}^* \leftarrow \arg\max_{i \in I} \log \rho_i \cdot \phi_i$
      $m \leftarrow 0$
  Otherwise, selects $i_t{}^*$ randomly
  Receives the reward $r_{u,i^*(t)}$
  If $r_{u,i^*(t)} = 0$: $m \leftarrow m + 1$
  Updates $\Sigma_{u,t} \leftarrow \Sigma_{u,t} + q_{i^*(t)} \cdot q_{i^*(t)}^\top$
  Updates $\mathbf{X} \leftarrow \mathbf{X} + r_{u,i^*(t)} \cdot q_{i^*(t)}$

Algorithm 2 CONTEXTUAL *LinUCB*

Require: features $Q = \{q_1, ..., q_n\}$ from SVD, popularity ρ, entropy ϕ, the value α to manage the item's uncertainty, threshold $\mathcal{T} > 0$

$\mathbf{X} \leftarrow BFGS_{\mathbf{x}} \sum_i (\log \rho_i \cdot \phi_i - \mathbf{x}^\top Q_i)^2$
$\Sigma_{u,t} \leftarrow I_d$
$m \leftarrow 0$
for $t = 1, 2, ..., T$ **do**
  Estimates $p_{u,t} \leftarrow \Sigma_{u,t}^{-1} \cdot \mathbf{X}$
  If $m < \mathcal{T}$:
    $i_t{}^* \leftarrow \arg\max_{i \in I \setminus R} p_{u,t}^\top q_i + \alpha ||q_i||_{2,\Sigma_{u,t}}$
  where $||q_i||_{2,\Sigma_{u,t}} = \sqrt{q_i^\top \Sigma_{u,t} q_i}$
  Otherwise:
    $i_t{}^* \leftarrow \arg\max_{i \in I} \log \rho_i \cdot \phi_i$
    $m \leftarrow 0$
  Receives the reward $r_{u,i^*(t)}$
  If $r_{u,i^*(t)} = 0$: $m \leftarrow m + 1$
  Updates $\Sigma_{u,t} \leftarrow \Sigma_{u,t} + q_{i^*(t)} \cdot q_{i^*(t)}^\top$
  Updates $\mathbf{X} \leftarrow \mathbf{X} + r_{u,i^*(t)} \cdot q_{i^*(t)}$

Algorithm 3 CONTEXTUAL *PTS*

Require: features Q and P from PMF, σ, σ_P, σ_Q, K particles, popularity ρ, entropy ϕ, and the threshold $\mathcal{T} > 0$

$\mathbf{X} \leftarrow BFGS_{\mathbf{x}} \sum_i (\log \rho_i \cdot \phi_i - \mathbf{x}^\top Q_i)^2$
Initialize the particles: $[d_{X_u} \leftarrow \mathbf{X}] \forall K$
$m \leftarrow 0$
for $t = 1, 2, ..., T$ **do**
  $d' \sim d_{\mathbf{w}}$
  $\bar{Q} \leftarrow d'_Q$
  $\bar{p}_u \sim P(p_u | \bar{Q}, d'_{\sigma_P}, \sigma, r^o_{1:t-1})$
  If $m < \mathcal{T}$:
    $i_t{}^* \leftarrow \arg\max_{i \in I \setminus R} \bar{p}_u \bar{q}_i$
  Otherwise:
    $i_t{}^* \leftarrow \arg\max_{i \in I} \log \rho_i \cdot \phi_i$
    $m \leftarrow 0$
  Receives the reward $r_{u,i^*(t)}$
  If $r_{u,i^*(t)} = 0$: $m \leftarrow m + 1$
  $r^o_t \leftarrow (u, i_t{}^*, r_{u,i^*(t)})$
  Updates d based on [17]

4/ RESULTADOS

EXPERIMENTACIÓN

Q1: *Are the modified versions of the bandit algorithms statistically superior to the original ones?*

Q2: *Has the improvement of these modified versions been caused by the usual popularity bias of offline datasets?*

Q3: *What is the effect of addressing uncertainty scenarios compared to existing state-of-the-art baselines?*

Datasets	# Users	# Items	Sparsity
Netflix	10,000	17,372	98.67%
GoodBooks	53,423	10,000	98.88%
Yahoo Music R1	10,000	13,214	99.22%

Table 1: An overview of the datasets applied in this work.

Dataset	Netflix 10k					Good Books					Yahoo Music 10k				
Measure	Hits					Hits					Hits				
T	5	10	20	50	100	5	10	20	50	100	5	10	20	50	100
Linear ϵ-Greedy	0.006	0.158	0.734	2.323	5.963	0.029	0.051	0.133	0.478	1.344	0.003	0.011	0.061	0.316	1.059
Linear ϵ-Greedy$_{t=0}$	0.546▲	1.161▲	2.636▲	7.378▲	15.475▲	0.418▲	0.952▲	2.014▲	5.616▲	11.185▲	0.936▲	1.924▲	4.005▲	9.746▲	17.186▲
Linear ϵ-Greedy$_{t\geq 0}$	**0.549**▲	**1.200**▲	**3.711**▲	**10.793**▲	**20.497**▲	**0.472**▲	**1.095**▲	**2.250**▲	**6.342**▲	**12.801**▲	**1.032**▲	**2.210**▲	**5.079**▲	**12.552**▲	**21.383**▲
PTS	1.116	2.300	4.299	7.560	12.037	0.870	2.168	4.428	8.035	12.455	**1.589**	**3.329**	**6.704**	14.363	19.634
PTS$_{t=0}$	0.619 ▼	1.508 ▼	3.554 ▼	9.069 ▲	15.606 ▲	**1.016**▲	**2.390**▲	**4.536**▲	8.267▲	12.547▲	1.465 ▼	3.183 ▼	6.644 ▼	14.623▲	20.203▲
PTS$_{t\geq 0}$	**1.364**▲	**2.772**▲	**4.830**▲	**10.358**▲	**15.784**▲	1.000▲	2.320▲	4.468●	**8.831**▲	**13.409**▲	1.500 ▼	3.239 ▼	6.556 ▼	**14.698**▲	**20.218**▲
LinUCB	0.452	0.739	3.429	12.469	23.030	0.969	**2.600**	5.291	10.270	15.809	1.399	**3.530**	**7.638**	16.751	25.477
LinUCB$_{t=0}$	**0.996**▲	**1.514**▲	**5.252**▲	14.534▲	24.905▲	**1.174**▲	1.893▼	**5.435**▲	**11.252**▲	**17.522**▲	**1.409**●	2.336▼	7.250▼	**17.922**▲	27.260▲
LinUCB$_{t\geq 0}$	**0.996**▲	**1.514**▲	**5.252**▲	**14.536**▲	**24.914**▲	**1.174**▲	1.893▼	5.425▲	11.241▲	17.497▲	**1.409**●	2.336▼	7.242▼	17.910▲	**27.334**▲

EXPERIMENTACIÓN

Q1: *Are the modified versions of the bandit algorithms statistically superior to the original ones?*

Q2: *Has the improvement of these modified versions been caused by the usual popularity bias of offline datasets?*

Q3: *What is the effect of addressing uncertainty scenarios compared to existing state-of-the-art baselines?*

Synthetic Dataset	
# Users	943
# Items	1,000
Sparsity	93.75%
Rating mean	3.25 (std. 1.27)
Avg. ratings per user	62 (std. 60)
Avg. ratings per item	59 (std. 26)

Table 3: Synthetic Dataset statistics.

Dataset	Synthetic Dataset								
Measure	Estimated policy value			95.0% CI (lower) – 95.0% CI (upper)			Relative policy value		
Estimators	IPS	DM	DR	IPS	DM	DR	IPS	DM	DR
Random	0.656	3.253	3.245	0.000 - 1.791	3.226 - 3.281	3.214 - 3.276	0.199	0.990	0.987
Most Popular	3.643	3.394	3.438	2.113 - 5.476	3.370 - 3.420	3.373 - 3.526	1.109	1.033	1.046
Linear ϵ-Greedy$_{t=0}$	4.461	3.885	3.848	1.954 - 7.451	3.858 - 3.913	3.79 - 3.895	**1.358**	**1.182**	**1.171**
Linear ϵ-Greedy$_{t\geq 0}$	4.714	3.843	4.026	2.44 - 7.151	3.815 - 3.871	3.798 - 4.307	**1.435**	**1.170**	**1.225**
PTS$_{t=0}$	3.682	4.373	4.365	1.312 - 6.712	4.349 - 4.397	4.339 - 4.393	**1.121**	**1.331**	**1.329**
PTS$_{t\geq 0}$	5.38	4.184	4.173	2.784 - 8.092	4.158 - 4.208	4.136 - 4.215	**1.638**	**1.273**	**1.270**
LinUCB$_{t=0}$	3.541	4.613	4.602	0.987 - 7.144	4.604 - 4.621	4.549 - 4.651	1.078	**1.404**	**1.401**
LinUCB$_{t\geq 0}$	3.856	4.606	4.593	1.12 - 7.505	4.597 - 4.615	4.54 - 4.645	**1.174**	**1.402**	**1.398**

EXPERIMENTACIÓN

Q1: *Are the modified versions of the bandit algorithms statistically superior to the original ones?*

Q2: *Has the improvement of these modified versions been caused by the usual popularity bias of offline datasets?*

Q3: *What is the effect of addressing uncertainty scenarios compared to existing state-of-the-art baselines?*

(1) **Direct Method (DM)**: use all imputed ratings for items selected by the known policy.

$$\hat{V}_{DM}(\pi_e; D, \hat{r}) := \frac{1}{n} \sum_{i=1}^{n} \sum_{a \in \mathcal{A}} \pi_e(a|x_a) \hat{r}(x_i, a)$$

(2) **Inverse Propensity Score (IPS)**: weight the current policy with the known value of the original policy π_0 used to create the dataset.

$$\hat{V}_{IPS}(\pi_e; D) := \frac{1}{n} \sum_{i=1}^{n} \frac{\pi_e(a|x_a)}{\pi_0(a_i|x_i)} \cdot r_i$$

(3) **Doubly Robust (DR)**: combine both DM and IPS to reduce the variance and work well with small samples.

$$\hat{V}_{DR}(\pi_e; D, \hat{r}) := \hat{V}_{DM}(\pi_e; D, \hat{r}) + \frac{1}{n} \sum_{i=1}^{n} \frac{\pi_e(a|x_a)}{\pi_0(a_i|x_i)} \cdot (r_i - \hat{r}(x_i, a_i))$$

EXPERIMENTACIÓN

Q1: *Are the modified versions of the bandit algorithms statistically superior to the original ones?*

Q2: *Has the improvement of these modified versions been caused by the usual popularity bias of offline datasets?*

Q3: *What is the effect of addressing uncertainty scenarios compared to existing state-of-the-art baselines?*

Dataset	Synthetic Dataset								
Measure	Estimated policy value			95.0% CI (lower) – 95.0% CI (upper)			Relative policy value		
Estimators	IPS	DM	DR	IPS	DM	DR	IPS	DM	DR
Random	0.656	3.253	3.245	0.000 - 1.791	3.226 - 3.281	3.214 - 3.276	0.199	0.990	0.987
Most Popular	3.643	3.394	3.438	2.113 - 5.476	3.370 - 3.420	3.373 - 3.526	1.109	1.033	1.046
Linear ϵ-Greedy$_{t=0}$	4.461	3.885	3.848	1.954 - 7.451	3.858 - 3.913	3.79 - 3.895	**1.358**	**1.182**	**1.171**
Linear ϵ-Greedy$_{t\geq 0}$	4.714	3.843	4.026	2.44 - 7.151	3.815 - 3.871	3.798 - 4.307	**1.435**	**1.170**	**1.225**
PTS$_{t=0}$	3.682	4.373	4.365	1.312 - 6.712	4.349 - 4.397	4.339 - 4.393	**1.121**	**1.331**	**1.329**
PTS$_{t\geq 0}$	5.38	4.184	4.173	2.784 - 8.092	4.158 - 4.208	4.136 - 4.215	**1.638**	**1.273**	**1.270**
LinUCB$_{t=0}$	3.541	4.613	4.602	0.987 - 7.144	4.604 - 4.621	4.549 - 4.651	1.078	**1.404**	**1.401**
LinUCB$_{t\geq 0}$	3.856	4.606	4.593	1.12 - 7.505	4.597 - 4.615	4.54 - 4.645	**1.174**	**1.402**	**1.398**

EXPERIMENTACIÓN

Q1: *Are the modified versions of the bandit algorithms statistically superior to the original ones?*

Q2: *Has the improvement of these modified versions been caused by the usual popularity bias of offline datasets?*

Q3: *What is the effect of addressing uncertainty scenarios compared to existing state-of-the-art baselines?*

EXPERIMENTACIÓN

Datasets	# Users	# Items	Sparsity
Netflix	10,000	17,372	98.67%
GoodBooks	53,423	10,000	98.88%
Yahoo Music R1	10,000	13,214	99.22%

Table 1: An overview of the datasets applied in this work.

Dataset	Netflix 10k					Good Books					Yahoo Music 10k				
Measure	Hits					Hits					Hits				
T	5	10	20	50	100	5	10	20	50	100	5	10	20	50	100
Random	0.029	0.057	0.104	0.247	0.484	0.039	0.077	0.155	0.375	0.757	0.015	0.032	0.072	0.192	0.390
Popular	**1.581▲**	2.771●	4.915	10.203	17.776	1.238	2.231	4.246	7.919	11.993	1.593	2.902	5.080	10.581	17.432
ϵ-Greedy	0.664	1.297	2.393	5.369	9.874	0.445	0.807	1.422	3.090	5.659	0.589	1.429	3.100	7.404	13.307
UCB	0.558	1.176	2.274	5.222	9.660	0.404	0.742	1.342	2.990	5.579	0.514	1.358	3.018	7.330	13.277
TS	1.006	1.944	3.535	7.502	12.981	0.792	1.393	2.354	4.605	7.297	0.957	1.907	3.697	8.356	14.720
Linear UCB	0.695	1.770	4.189	11.377	21.380	0.435	1.020	2.792	7.589	13.713	**1.617▲**	**3.333▲**	6.732	15.524	25.134
GLM-UCB	0.643	1.344	3.990	11.627	21.701	0.333	0.719	1.551	7.030	13.284	0.975	1.704	4.975	14.190	24.455
Linear ϵ-Greedy	0.006	0.158	0.734	2.323	5.963	0.029	0.051	0.133	0.478	1.344	0.003	0.011	0.061	0.316	1.059
LinUCB	0.452	0.739	3.429	12.469	23.030	0.651	0.889	3.572	9.541	15.758	0.883	1.157	1.521	2.696	4.054
PTS	1.116	2.300	4.299	7.560	12.037	0.870	2.168	4.428	8.035	12.455	1.589	3.329	6.704	14.363	19.634
NICF	1.395	2.542	4.700	9.350	14.146	**1.685▲**	**3.011▲**	4.644	7.605	9.946	1.587	3.088	5.670	11.204	14.422
Cluster Bandit	0.571	1.230	3.132	7.420	13.882	0.944	1.792	4.001	7.577	11.835	0.996	2.428	4.704	10.095	16.777
ICTRTS	0.016	0.052	0.339	2.148	5.091	0.330	1.054	2.723	6.897	11.251	0.011	0.138	1.167	6.149	13.446
Linear ϵ-Greedy$_{t\geq 0}$	0.549	1.200	3.711	10.793	20.497	0.472	1.095	2.250	6.342	12.801	1.032	2.210	5.079	12.552	21.383
PTS$_{t\geq 0}$	1.364	**2.772●**	4.830	10.358	15.784	1.000	2.320	4.468	8.831	13.409	1.500	3.239	6.556	14.698	20.218
LinUCB$_{t\geq 0}$	0.996	1.514	**5.252▲**	**14.536▲**	**24.914▲**	1.174	1.893	**5.425▲**	**11.241▲**	**17.497▲**	1.409	2.336	**7.242▲**	**17.910▲**	**27.334▲**

Q1: *Are the modified versions of the bandit algorithms statistically superior to the original ones?*

Q2: *Has the improvement of these modified versions been caused by the usual popularity bias of offline datasets?*

Q3: *What is the effect of addressing uncertainty scenarios compared to existing state-of-the-art baselines?*

(1) The Linear ϵ-Greedy algorithm has been transformed into a highly competitive option. Before our modifications, the Linear ϵ-Greedy was one of the worst models, taking many interactions ($T \gg 100$) to learn user preferences. After it, this simple algorithm outperforms other approaches that take so much time to execute, like the Cluster-Bandit; and approaches that require so much effort to calibrate their parameters, like the NICF, ICTRTS, and PTS.

(2) The modified PTS model becomes more competitive with the other algorithms in the long run. With the addition of entropy to the already biased popularity model, it learns more about user preferences and thus maximizes their experience. However, as the model was already so biased by the popularity distribution, its results are not as good as the other modified versions.

(3) Especially, the modified LinUCB outperforms all of the state-of-the-art algorithms in the long run by achieving a bigger cumulative reward. After mitigating both scenarios of uncertainty, this method achieves better results even than strategies developed with similar assumptions, like Linear UCB and GLM-UCB.

EXPERIMENTACIÓN

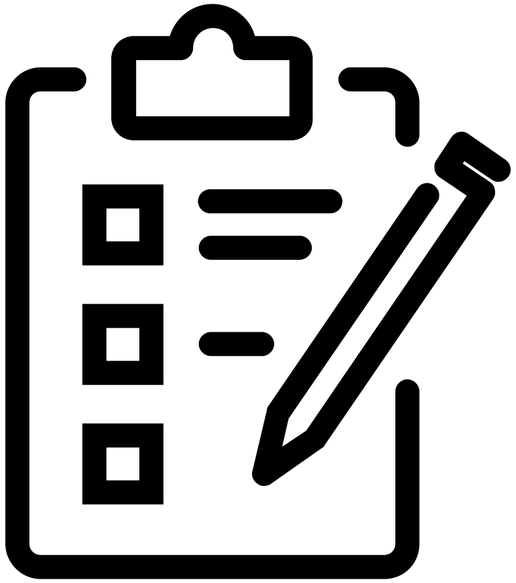
Q1: *Are the modified versions of the bandit algorithms statistically superior to the original ones?*

Q2: *Has the improvement of these modified versions been caused by the usual popularity bias of offline datasets?*

Q3: *What is the effect of addressing uncertainty scenarios compared to existing state-of-the-art baselines?*

5/ CONCLUSIONES Y HALLAZGOS

5/ CONCLUSIONES Y HALLAZGOS



- Mitigar la incertidumbre sobre las preferencias del usuario mejora la calidad de las recomendaciones a largo plazo.
- El AA permite a los modelos aprender de manera más efectiva y eficiente en escenarios de incertidumbre.
- Los algoritmos modificados, incluso los más simples como Linear ϵ-Greedy, se vuelven altamente competitivos al incorporar AA.

Trabajo futuro: Evaluación en línea, nuevas estrategias de AA, identificación de otros escenarios de incertidumbre.

MUCHAS GRACIAS

REFERENCIAS

SILVA, N., SILVA, T., HOTT, H., RIBEIRO, Y., PEREIRA, A., & ROCHA, L. (2023). EXPLORING SCENARIOS OF UNCERTAINTY ABOUT THE USERS' PREFERENCES IN INTERACTIVE RECOMMENDATION SYSTEMS. PROCEEDINGS OF THE 46TH INTERNATIONAL ACM SIGIR CONFERENCE ON RESEARCH AND DEVELOPMENT IN INFORMATION RETRIEVAL (SIGIR '23), JULY 23–27, 2023, TAIPEI, TAIWAN. HTTPS://DOI.ORG/10.1145/3539618.3591684