RED: Recommendations Encouraging Diversity

Project check-in

25.03.2025

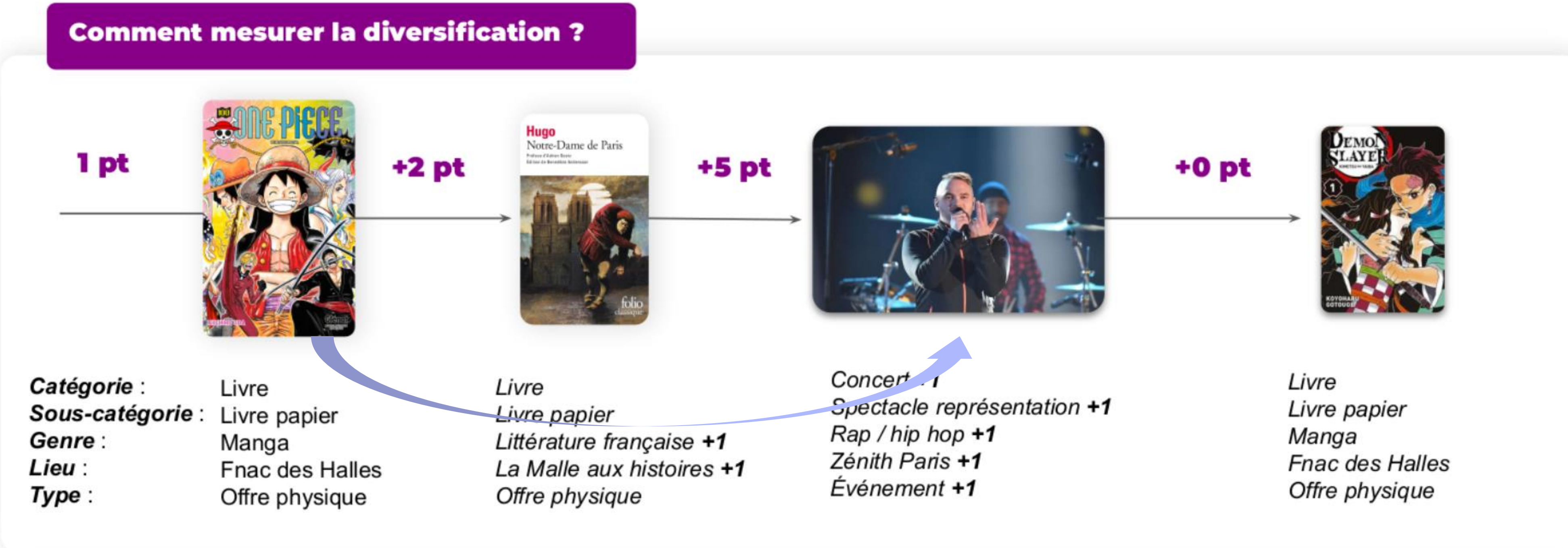
Clemence Reda

Background

1. Motivation: Pass Culture & drug repurposing
2. Objectives of the RED project
3. Data and diversity metrics
4. Intro to Multi-armed bandits
5. Intro to Determinantal Point Processes

Pass Culture a phone app for French teens (<20yr) to browse and book cultural goods nearby with credits.

Diversification points obtained for each new category / subcategory / genre / location / type (a bit like set cover; achievement score); those are not visible to the user



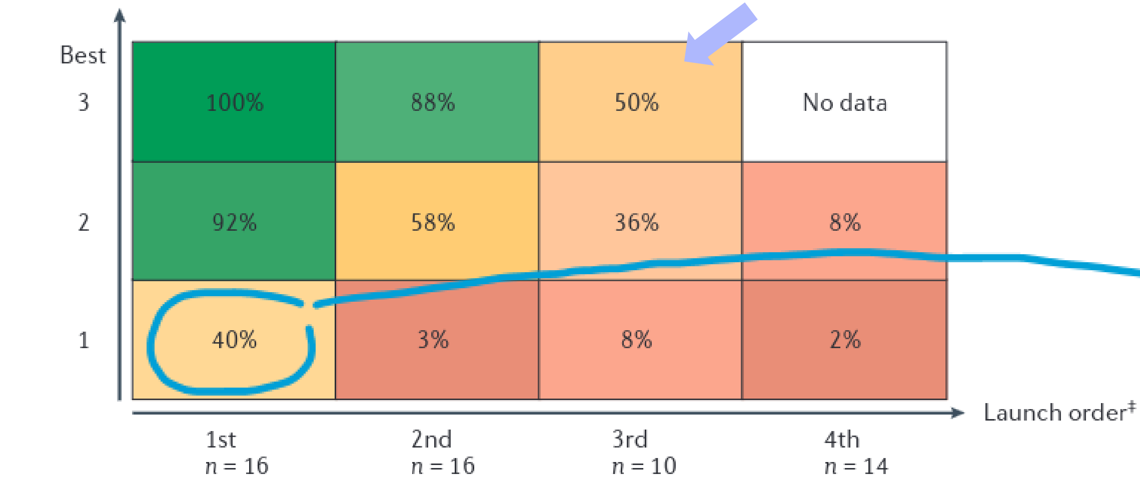
Courtesy of Jill-Jenn Vie (Inria SODA).

Drug repurposing First-in-class versus Best-in-class

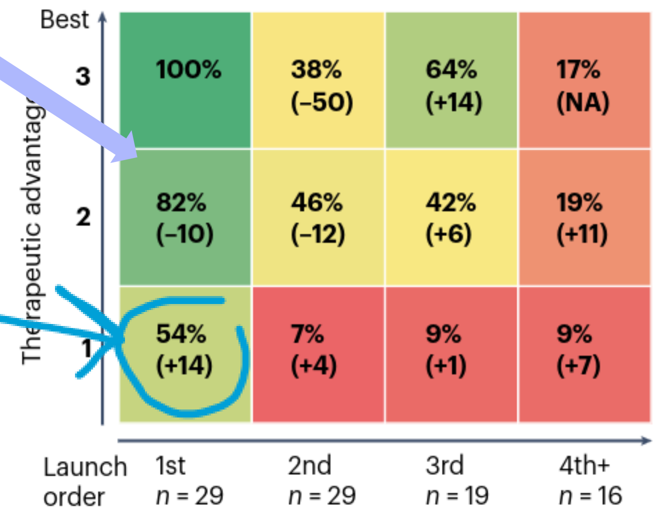
First launched in that mechanistic class — Highest therapeutic advantage [...]

In 2013 [1]

Trend in market value (aka sales)



In 2023 [2]



"The data indicated that it is slightly better to be first than to be best"

"[...] products that are first-to-launch increasingly tend to perform better [...]"

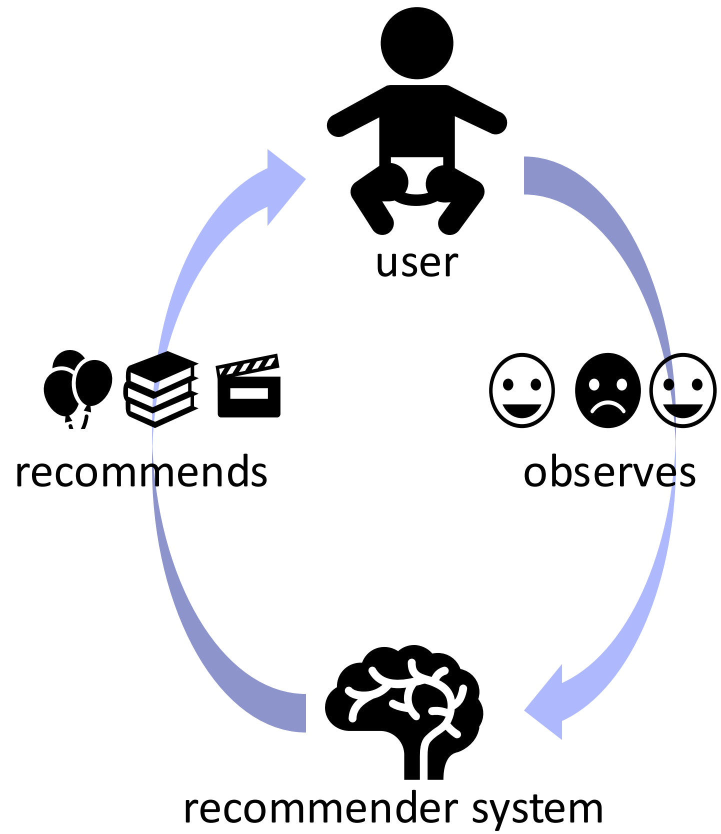
[1] Schulze, U., & Ringel, M. (2013). What matters most in commercial success: first-in-class or best-in-class?. *Nature Reviews Drug Discovery*, *12*(6), 419-420.
[2] Spring, L., Demuren, K., Ringel, M., & Wu, J. (2023). First-in-class versus best-in-class: an update for new market dynamics. *Nat Rev Drug Discov*, *22*(7), 531-532.

Objective of RED to design recommender systems for *personalized* *good* and *diverse* items.

taking into account the user's interests/item history

suggesting items with a high probability of positive feedback from the user

somewhat "controllable" recommendations: *e.g.*, out of the user's comfort zone



State-of-the-art some keywords

- Serendipity, diversity: metrics (see next slides)
- Contextual recommender systems: (logistic/cascading) multi-armed bandits (online), collaborative filtering (offline)

Scientific challenges

- Incorporate a tradeoff between quality and diversity
- Versatile enough to be applicable in all use cases

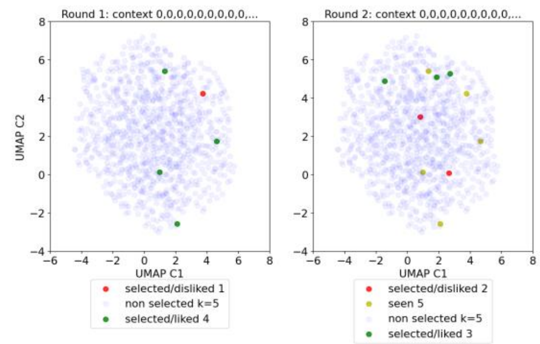
Data and diversity metrics different levels for development (D) and production (P)

Pass Culture data (P) Collection of events and item embeddings from the phone app for 6 months
delta: diversity reward: diversity+booking

```
df[['user', 'item', 'context', 'event_name', 'category_id', 'delta', 'booking', 'reward']]
```

	user	item	context	event_name	category_id	delta	booking	reward
204770	17061	123583	0	HasAddedOfferToFavorites	5	1	0	0
670169	9374	47750	0	BookingConfirmation	7	1	1	2
893131	39047	119265	4	ConsultOffer	8	1	0	0
518769	35989	27319	128	BookingConfirmation	7	0	1	1
482951	37074	9258	0	ConsultOffer	7	1	0	0

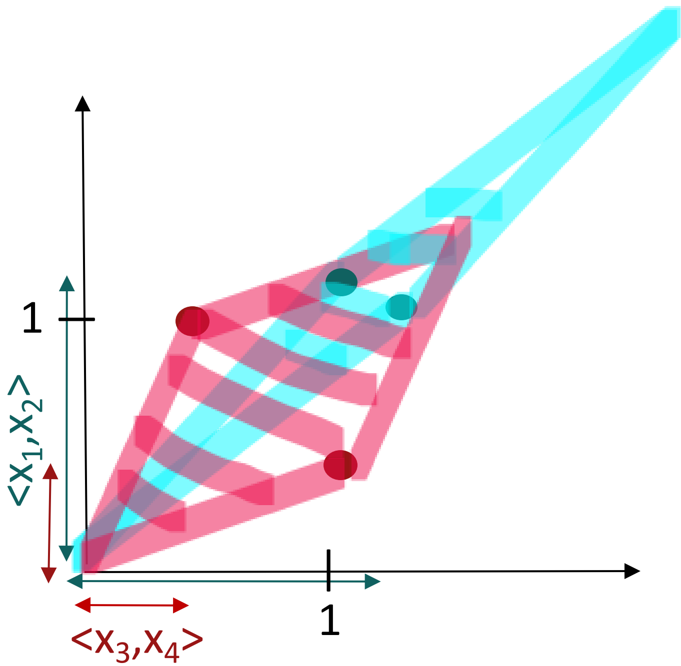
Synthetic data (D) Generate item embeddings x_i and θ at random (Gaussian) to get trajectories (starting with context=0) where reward ~ Ber($\sigma(\underbrace{z_{iu}}_{\text{mixture of context and item embedding}}{}^T\theta)$) for user u, item i



Pseudo-real MovieLens [1] data (D)
Item embeddings: one-hot encoding of genre keywords, fit θ using a shallow MLP **[2]**

[1] https://movielens.org/
[2] Papini, M., Tirinzoni, A., Restelli, M., Lazaric, A., & Pirotta, M. (2021). Leveraging good representations in linear contextual bandits. In *International Conference on Machine Learning* (pp. 8371-8380). PMLR.

Data and diversity metrics item-embedding-driven diversity metrics: (log)-determinant, ridge leverage score



$$K_{\{xi,xj\}} = \begin{pmatrix} 1 & <x_i,x_j> \\ <x_i,x_j> & 1 \end{pmatrix}$$

$$= 1 - <x_i,x_j>^2 \geq 0$$

$|K_{\{x1,x2\}}| \leq |K_{\{x3,x4\}}|$
where $<x_3,x_4> \leq <x_1,x_2>$ (cosine similarity)

(Item embedding) kernel K $K_{\{x,y\}} = <\Phi(x), \Phi(y)>$
e.g., linear kernel: Φ = Id, ~ similarity b/w items

(Log)-determinant of (a subset of) the kernel Volume in space occupied by a set S of items: $|K_S|$

Effective dimension of (a subset of) the kernel for a ridge factor λ and a set S of items:

$$d_{eff}(S) = Tr(K_S(K_S+\lambda I_d)^{-1})$$

is the sum of (ridge) leverage scores for each $i \in S$

Drawn for all i $|x_i|^2=1$, K linear kernel in 2D
... but works for any # of dimension and # of points (and any symmetric kernel)

Data and diversity metrics item-embedding-driven diversity metrics: (log)-determinant, ridge leverage score

The ridge leverage score for i^{th} item in set S is

$$(K_S(K_S+\lambda I_d)^{-1})_{ii} \leq 1$$

and also the optimal value of the ridge regression problem **[1]**

$$\min_{\psi} \underbrace{|\psi B-b_i}_{\text{find a linear combination of rows of B}}|^2+\lambda|\psi|^2$$

where $BB^T = K_S$ and $B=[b_i]_{i\in S}$

(Item embedding) kernel K $K_{\{x,y\}} = \langle\Phi(x), \Phi(y)\rangle$

e.g., linear kernel: Φ = Id, ~ similarity b/w items

(Log)-determinant of (a subset of) the kernel Volume in space occupied by a set S of items: $|K_s|$

Effective dimension of (a subset of) the kernel for a ridge factor λ and a set S of items:

$$d_{eff}(S) = Tr(K_s(K_s+\lambda I_d)^{-1})$$

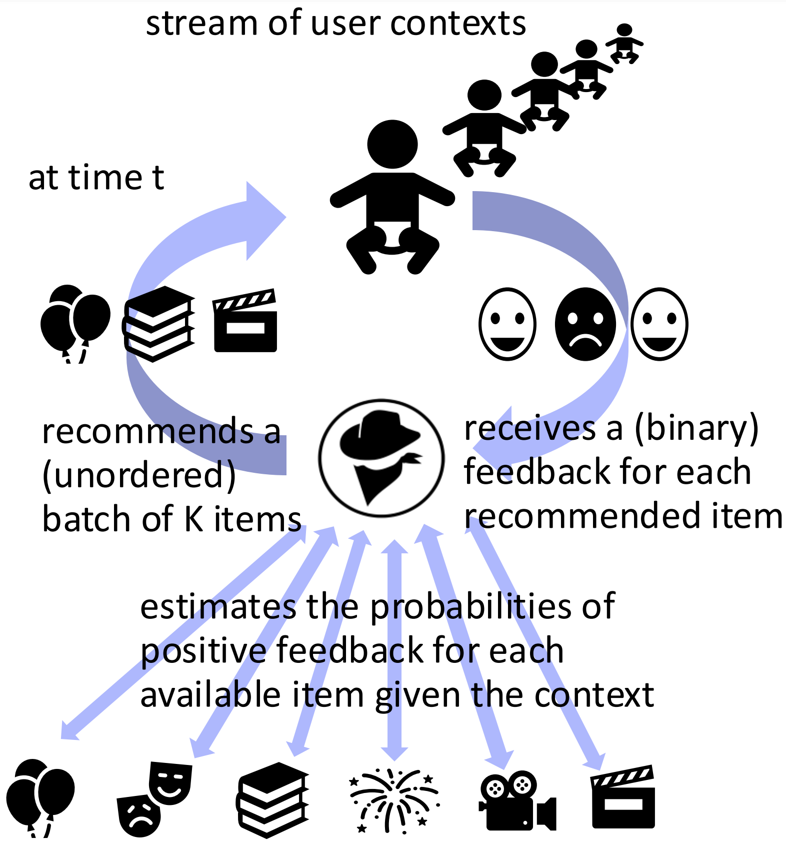
is the sum of (ridge) leverage scores for each $i\in S$

The higher the effective dimension, the higher the number of "degrees of freedom", "uniqueness of items" in the set

[1] Musco, C., & Musco, C. (2017). Recursive sampling for the nystrom method. *Advances in neural information processing systems*, *30*.

(Stochastic, contextual) multi-armed bandits (MABs)

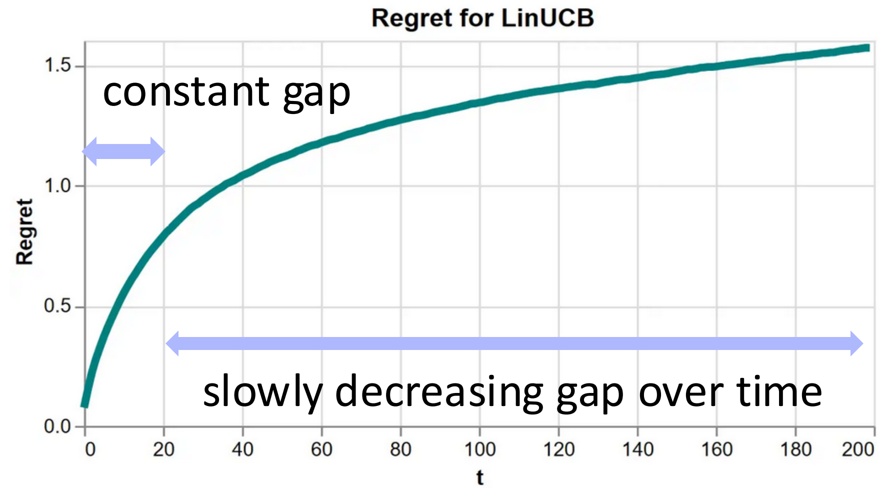
online recommender systems for reward maximization



Definition of a MAB A MAB is defined by its sampling rule: highest probability of positive feedback?

Reward max/regret min Performance is compared to a deterministic oracle with access to the true probs

Regret = Perf(Oracle) - Perf(Algo)



It is a vast research domain, feel free to check out

[1] Lattimore, T., & Szepesvári, C. (2020). *Bandit algorithms*. Cambridge University Press.

(Finite) determinantal Point Processes (DPPs) sampling diverse set of points in a data-driven fashion

Point Process A distribution over finite subsets of a (finite) set $X=\{x_1,x_2,\ldots,x_N\}$

Definition of a DPP* The probability of a subset is positively correlated to its diversity

$$\mathrm{Prob}(S \subseteq X) = |K_S|/|K+I_d| \text{ where } K \text{ is the kernel built on } \{x_1,x_2,\ldots,x_N\}$$

There are algorithms in $O(N^3)$ to sample k points out of N from a DPP **[1]**

Finding a subset S maximizing $\mathrm{Prob}(S \subseteq X)$ is NP-hard

If you want to test DPPs, check out the Python library DPPy

[1] Theorem 7 and Algorithm 18 in Hough, J. B., Krishnapur, M., Peres, Y., & Virág, B. (2006). Determinantal processes and independence.

It is (again) a vast research domain, feel free to check out

[2] Kulesza, A., & Taskar, B. (2012). Determinantal point processes for machine learning. *Foundations and Trends® in Machine Learning*, *5*(2–3), 123-286.

* That definition is actually that of a L-ensemble, which matches the definition of kernel given earlier.

(Finite) determinantal Point Processes (DPPs) leveraging the quality-diversity decomposition

Quality-decomposition Assume $K = Q\Phi^T\Phi Q$ where Φ is the item embedding matrix $|\Phi_i|=1$ and $Q = \mathrm{Diag}(q_1, q_2, \ldots, q_N)$ where $q_i \geq 0$ is the "quality" of item I

$$\mathrm{Prob}(S \subseteq X) \propto \Pi_{i \in S}\, q_i^2\, |\Phi_S^T\Phi_S|$$

Nyström approximation **[1-2]** If large dimension N of Φ: assume that K is of rank m << N, choose m "representative" points and apply a SVD to find an approximation of K depending only on smaller matrices

Greedy algorithm for maximizing set-function **[3]** Under some conditions on f*, $\mathrm{argmax}_S\, f(S)$ is built greedily by setting $S^\star = \{s_1, s_2, \ldots, s_K\}$ where $s_k = \mathrm{argmax}_i\, f(\{s_1, \ldots, s_{k-1}\} \cup \{i\})$ with a "small" approximation rate

[1] Drineas, P., & Mahoney, M. W. (2005). Approximating a gram matrix for improved kernel-based learning. In *International Conference on Computational Learning Theory* (pp. 323-337). Berlin, Heidelberg: Springer Berlin Heidelberg.
[2] https://andrewcharlesjones.github.io/journal/nystrom-approximation.html
* monotone, submodular
[3] Kempe, D., Kleinberg, J., & Tardos, É. (2003). Maximizing the spread of influence through a social network. In *Proceedings of the ninth ACM SIGKDD international conference on Knowledge discovery and data mining* (pp. 137-146).

My progress on the project

1. Formalization of the problem
2. Implementation of baselines
3. First approaches

Formalization of the problem gives a parametrization of the target function and how to learn from offline data

😮‍💨 Translating into a doable ML problem the list of requirements and ideas from real-life applications

Key ("new") problems:

1. Diversity without hurting quality (too much)
2. Missing feedback
3. Large # of items
4. Do not recommend a visited item again (++ contexts and item embeddings)

Input

Items normalized embeddings of size d
Contextual setting
At time t, a (new or old) user appears with her context/history C_t on all items
$C_t^i > 0$ if i was liked $=0$ if not visited
$C_t^i < 0$ if i was disliked

Output

Recommend K items $\{r^1_t, r^2_t, \ldots, r^K_t\}$
Observations User outputs immediately her feedback
$Y_t^i > 0$ if i was liked $Y_t^i=0$ if i was not visited
$Y_t^i < 0$ if i was disliked
Intrabatch diversity diversity among $r^1_t, r^2_t, \ldots, r^K_t$
Interbatch diversity diversity across t's

Formalization of the problem gives a parametrization of the target function and how to learn from offline data

Assumptions:

- Fixed random proba p^{visit} of visiting a recommended item
- Reward (feedback) for visited item i and context C_t is ~ $Ber(\sigma(z_{it}^T\theta^\star))$ where $z_{it} = [C_t, x_i]$
- Diversity is deterministic with a function $f^{div}(S_1, S_2)$ (*e.g.*, log det or effective dimension)
- Feedbacks are iid

$$\mathrm{Prob}(Y_t^i = y) = \begin{cases} 1-p^{visit} & \text{if } y=0 \\ p^{visit} \times \sigma(z_{it}^T\theta^\star) & \text{if } y = +1 \\ p^{visit} \times (1-\sigma(z_{it}^T\theta^\star)) & \text{if } y = -1 \end{cases}$$

Input

Items normalized embeddings of size d
Contextual setting
At time t, a (new or old) user appears with her context/history C_t on all items
$C_t^i > 0$ if i was liked =0 if not visited
$C_t^i < 0$ if i was disliked

→

Output

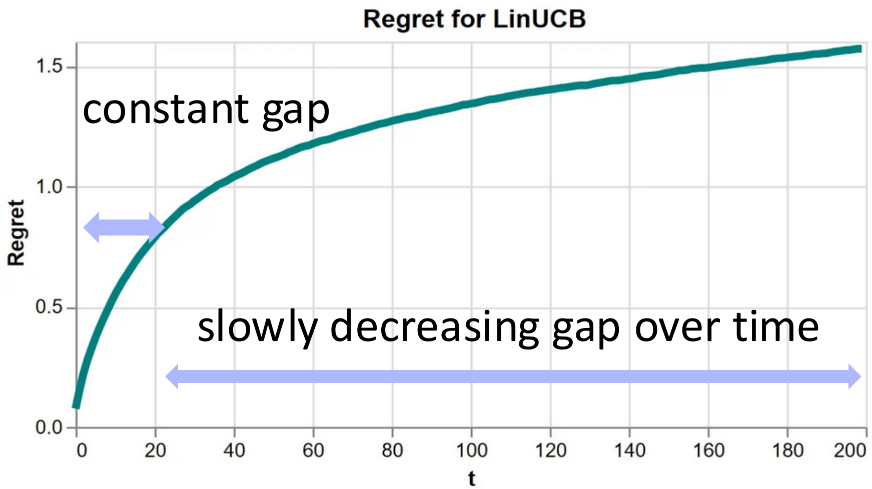
Recommend K items $\{r^1_t, r^2_t, \ldots, r^K_t\}$
Observations User outputs immediately her feedback
$Y_t^i > 0$ if i was liked $Y_t^i = 0$ if i was not visited
$Y_t^i < 0$ if i was disliked
Intrabatch diversity diversity among $r^1_t, r^2_t, \ldots, r^K_t$
Interbatch diversity diversity across t's

From now we assume that $p^{visit} = 1$ to make things easier

Formalization of the problem gives a parametrization of the target function and how to learn from offline data

Regret minimization Performance is compared to a deterministic oracle with access to the true probs (here, $\theta^\star$)

$$\text{Regret} = \text{Perf(Oracle)} - \text{Perf(Algo)}$$



Here we have three kinds of regrets!*

Reward Oracle Given $\theta^\star$, recommend

$$\pi^{rew}(C_t) = \text{top K of } \{\sigma(z_{it}^T\theta^\star)\}_i$$

Interbatch Diversity Oracle Recommend

$$\pi^{e\text{-}div}(C_t) = \text{top K-subset S for } f^{div}(S, C_t)$$

Intrabatch Diversity Oracle Recommend

$$\pi^{a\text{-}div}(C_t) = \text{top K-subset S for } f^{div}(S, \emptyset)$$

Computing those oracles is NP-hard 🙃 this is the link to DPPs

* You could scalarize everything (e.g., linear combination) but... you might not find the full Pareto front:

[1] Drugan, M. M., & Nowe, A. (2013). Designing multi-objective multi-armed bandits algorithms: A study. In *The 2013 international joint conference on neural networks (IJCNN)* (pp. 1-8). IEEE.

Relevant baselines as the problem is related to logistic bandits and diversity/coverage literature

Adaptations of → Restriction to non-visited items

- Ɛ-greedy: (incrementally) regress on θ, consider the top K items for $\sigma(z_{it}^T\theta)$
Flip a coin for each and replace by random item with proba Ɛ

- "Logistic Regression": regress on θ, apply Greedy on $q_i = \sigma(z_{it}^T\theta)$, $\Phi_i = x_i$

- LogisticUCB1 **[1]**: regress on θ, recommend K items with highest UCBs
$\sigma(z_{it}^T\theta) \leq \mathrm{UCB}(i, t)$ for all i and t with high proba

- "LinOASM" **[2]**: regress on θ (for logistic models), apply Greedy on $q_i = \mathrm{UCB}(i, t)$, $\Phi_i = x_i$

[1] Faury, L., Abeille, M., Calauzènes, C., & Fercoq, O. (2020). Improved optimistic algorithms for logistic bandits. In *International Conference on Machine Learning* (pp. 3052-3060). PMLR.
[2] Gabillon, V., Kveton, B., Wen, Z., Eriksson, B., & Muthukrishnan, S. (2014). Large-scale optimistic adaptive submodularity. In *Proceedings of the AAAI Conference on Artificial Intelligence* (Vol. 28, No. 1).

First approaches given the formulation we proposed

Meta-algorithm

Tradeoff between diversity and reward

Receive context C_t

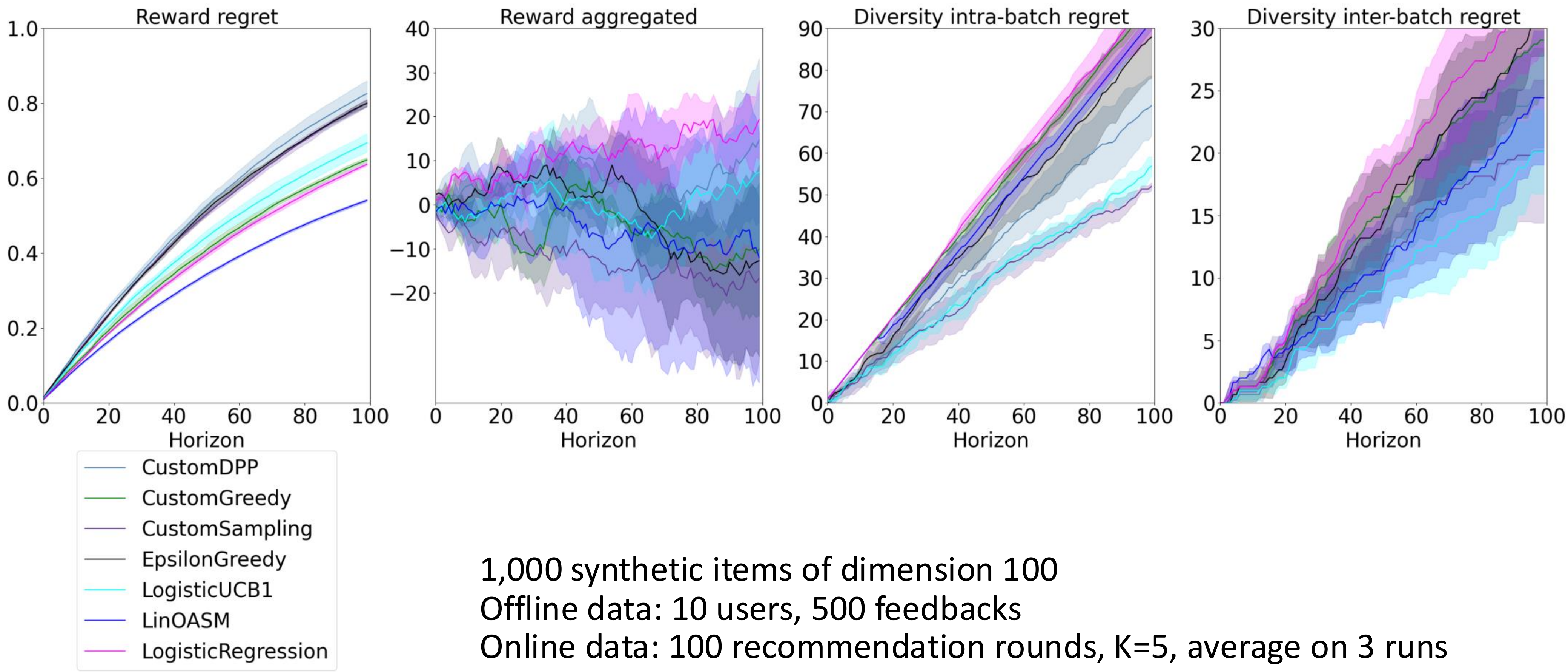
1. Compute q_i = ReLU$(\sigma(z_{it}{}^T\theta))^\alpha$, $\Phi_i = x_i$
2. Recommend $S_t \leftarrow$ sample(Q, Φ)
3. Update θ with Y_t

CustomDPP Apply a DPP for $K = Q\Phi^T\Phi Q$

CustomGreedy Apply Greedy algorithm

CustomSampling Sample at random M subsets, return maximizer

Experimental results on synthetic data sets



1,000 synthetic items of dimension 100
Offline data: 10 users, 500 feedbacks
Online data: 100 recommendation rounds, K=5, average on 3 runs

What happens next

1. See what happens on the MovieLens data set
2. Modify the estimator for θ when $p^{visit} < 1$
3. Make the implementation faster
4. Propose a final approach
5. Derive theoretical guarantees on (all) regrets