Diversified recommendations of cultural activities with personalized determinantal point processes

Carole Ibrahim Hiba Bederina Daniel Cuesta
Laurent Montier Cyrille Delabre Jill-Jênn Vie

Soda kick-off seminar, June 3, 2025

Pass Culture

Since 2019, the French government awards a fixed credit to 3 million young individuals (aged 15-20) to spend on $\sim$ 1 million possible activities (books, cinema, opera, etc.).

We want to:

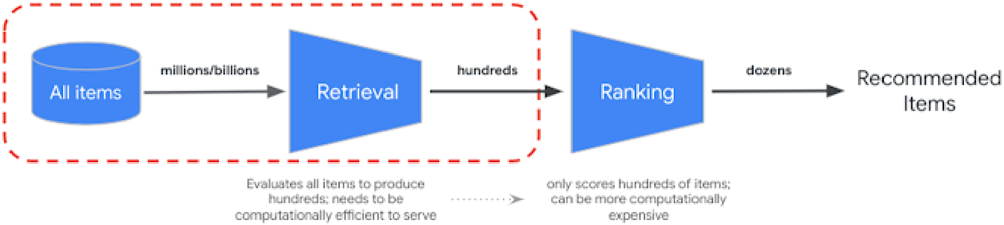
- increase youth participation in cultural activities
- broaden their cultural horizons: make them discover new things

How to model it? (1.5 year project)

Industrial recommender systems are vector databases

Among the million of offers, only 1500 are selected for ranking

Vector database: approximate nearest neighbor according to a query vector

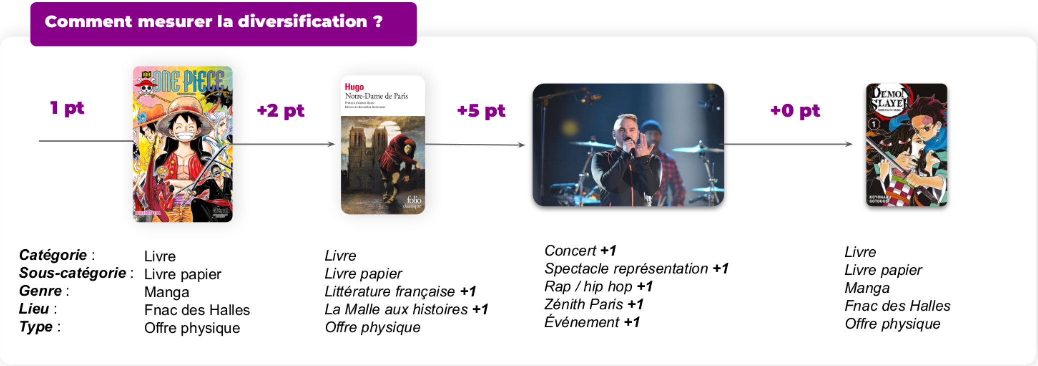


- One model for retrieval (two-tower model $\sim$ neural collaborative filtering)
- Another one for top K ranking (LightGBM; I also tried skrub)

Reward metrics (key performance indicators) of Pass Culture

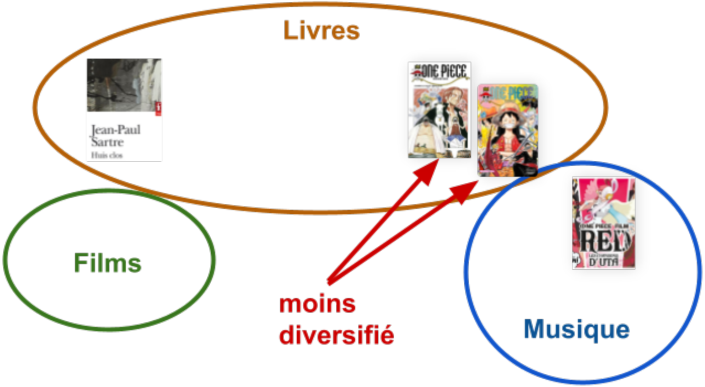
Relevance: click-through rate (booking rate)

Diversification points obtained for each new category / genre / location (increase in cultural diversity); those scores are not visible to the user, but for stakeholders

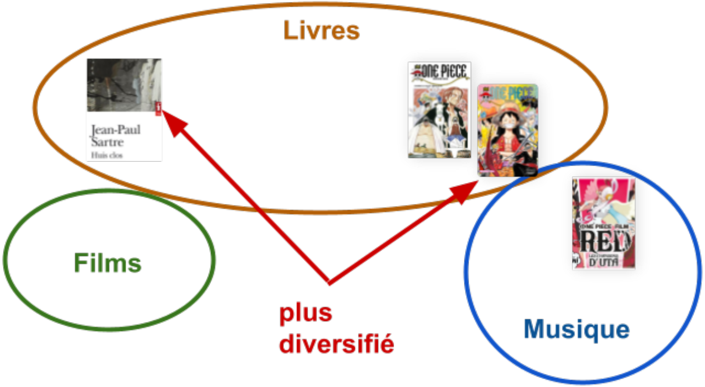


It somehow has limitations

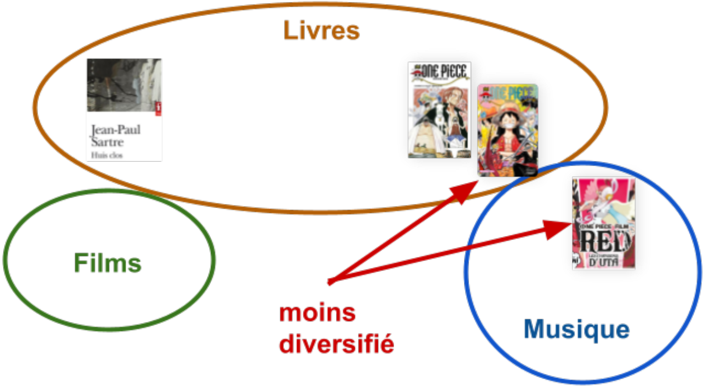
Parcours dans l'espace sémantique de la culture



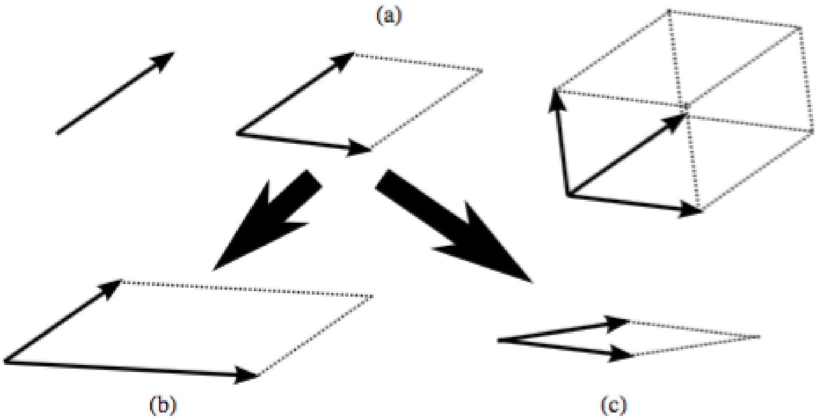
Parcours dans l'espace sémantique de la culture



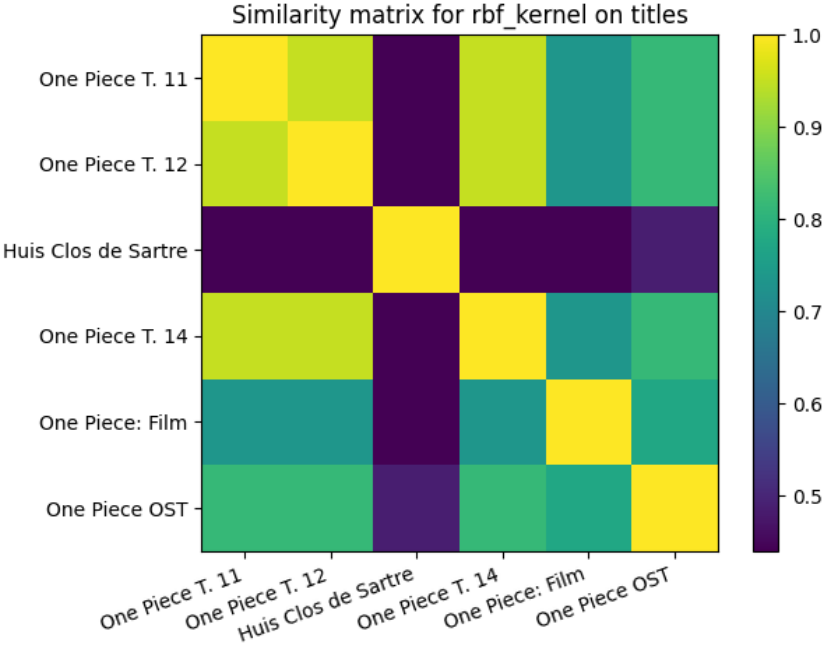
Parcours dans l’espace sémantique de la culture



Geometric modeling of diversity



- Determinant = square of volume of parallelotope of vectors
- Vectors that are not correlated increase the volume
- We want to sample items proportionally to diversity


Similarity matrix for rbf_kernel on titles
One Piece T. 11
One Piece T. 12
Huis Clos de Sartre
One Piece T. 14
One Piece: Film
One Piece OST
1.0
0.9
0.8
0.7
0.6
0.5

Quality-diversity decomposition for recommendation

- $q_i > 0$ is a possibly personalized measure of quality of item i for the current user
- ϕ_i is a unit semantic embedding of item i, $||\phi_i|| = 1$, used for diversity sampling

Similarity matrix $K = XX^T$ and $K_{ij} = x_i^T x_j$ can be decomposed as $q_i \phi_i^T \phi_j q_j$

Metrics of a set S for a user

1. Relevance, i.e. click-through rate

$$\frac{1}{|S|} \sum_{i \in S} q_i$$

2. Volume formed by set S

$$Vol(S)$$

3. Diversification is the increase in diversity

$$\Delta \simeq Vol(H \cup S) - Vol(H)$$

Our sampling objective

Sampling a set S proportional to $\det K_S$

$$\log \det K_S = \underbrace{\sum_{i \in S} \log q_i}_{\text{quality}} + 2 \underbrace{\log Vol(S)}_{\text{diversity}}$$

DPP

If we sample among n items
$K : n \times n$ similarity matrix on items (positive semi-definite)

P is a determinantal point process if sample Y verifies:

$$\forall A \subset \{1, \ldots, n\}, \quad P(A \subseteq Y) \propto det(K_A) = Vol(\{x_i\}_{i \in A})^2$$

where K_A has subset A of rows and columns.

There is a $O(nk^3)$ algorithm for sampling k items among n, at the cost of knowing its eigenvalues in $O(n^3)$, or $O(nd^2)$ for the linear kernel.

Example for sampling 3 points among 4

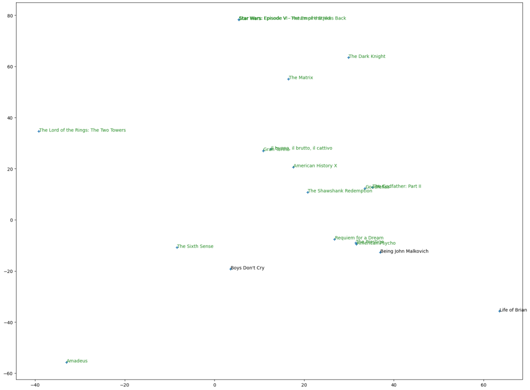
$$K = \begin{pmatrix} 1 & 2 & 3 & 4 \\ 2 & 5 & 6 & 7 \\ 3 & 6 & 8 & 9 \\ 4 & 7 & 9 & 1 \end{pmatrix}$$

$A = \{1, 2, 4\}$ will be included with probability proportional to

$$K_A = det \begin{pmatrix} 1 & 2 & 4 \\ 2 & 5 & 7 \\ 4 & 7 & 1 \end{pmatrix}$$

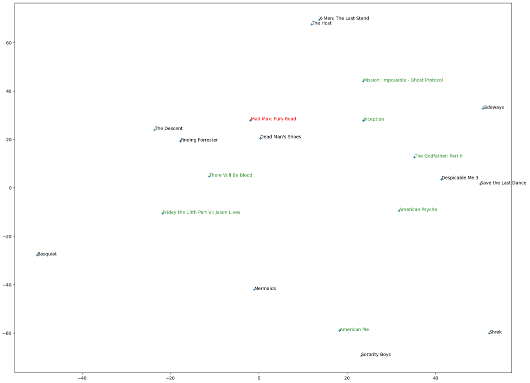
Compromise quality-diversity

SVD naive top K



Several Star Wars movies in the set

k-DPP



Evaluation

We conducted offline and online experiments (A/B/C test) on 400k users.

- Version A (baseline): recommender system
- Version B: DPP filter using personalized quality scores q_i
- Version C: DPP filter using $q_i = 1$

DPPs are implemented in DPPy by former colleague Guillaume Gautier at Inria Lille

Guillaume Gautier et al. "DPPy: DPP Sampling with Python". In: *Journal of Machine Learning Research* 20.180 (2019), pp. 1–7. URL: http://jmlr.org/papers/v20/19-179.html

Stochastic or deterministic?

We sample k-DPP proportionally to $\det K_S$

YouTube [2] computes instead the greedy max of $\operatorname*{argmax}_{S, |S| = k} \det K_S$

They happily reported "+0.5%" of increased user engagement (significant? ¯_(ツ)_/¯)

We hypothesize that a deterministic approach does not cover the catalogue well

Mark Wilhelm et al. "Practical diversified recommendations on youtube with determinantal point processes". In: *Proceedings of the 27th ACM International Conference on Information and Knowledge Management*. 2018, pp. 2165–2173

Results

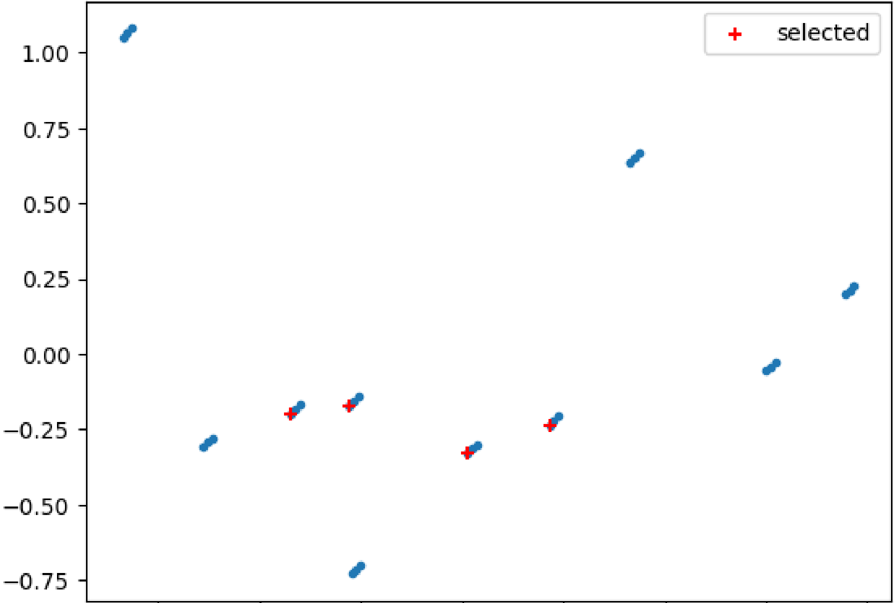
	Relevance	Volume ratio	Diversification
Model A	**0.525**	1	2.759
Model B	0.399	$\times$24.7	**3.404**
Model C	0.381	$\times$**28.8**	**3.482**

Table 1: Offline results comparing baseline (A) vs DPP-based recommenders (B and C).

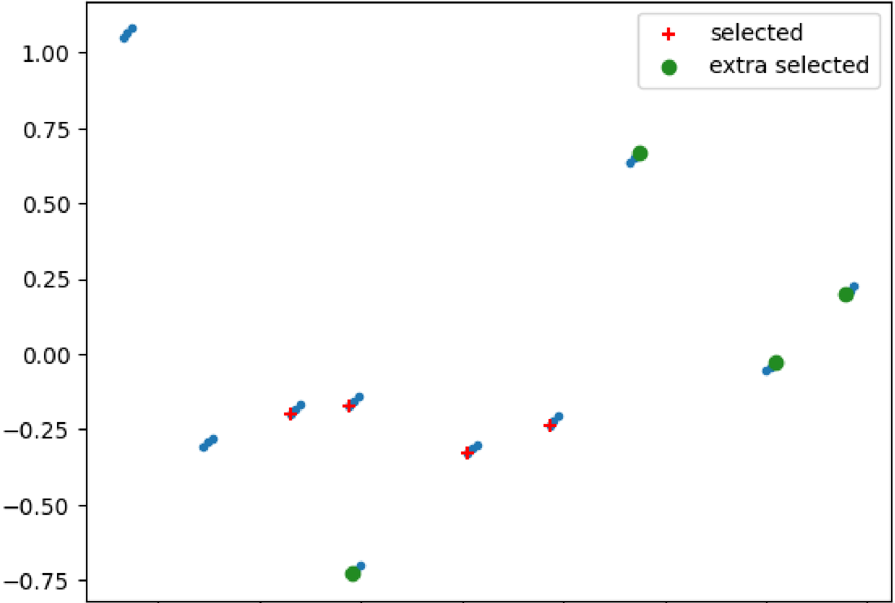
	Click rate	Volume ratio	Diversification
Group A	**0.54%**	1	3.132
Group B	0.34%*	$\times$12	**3.512***
Group C	0.29%*	$\times$**15.8**	**3.590***

Table 2: Online A/B/C test results. Values with * denote statistical significance ($p < 0.001$).

Conditional DPP for directly optimizing diversification



Conditional DPP for directly optimizing diversification



Thank you for your attention!

[1] Guillaume Gautier et al. "DPPy: DPP Sampling with Python". In: *Journal of Machine Learning Research* 20.180 (2019), pp. 1–7. URL: http://jmlr.org/papers/v20/19-179.html.
[2] Mark Wilhelm et al. "Practical diversified recommendations on youtube with determinantal point processes". In: *Proceedings of the 27th ACM International Conference on Information and Knowledge Management*. 2018, pp. 2165–2173.