Preference Elicitation in Mangaki: Is Your Taste Kinda Weird?

Jill-Jênn Vie

MANGAKI
LA RECOMMANDATION FRANCOPHONE D'ANIME ET MANGA

June 22, 2016

France is the 2nd manga consumer in the world

1. Japan - 500M volumes
2. France - 13M volumes
3. US - 9M volumes

Japan Expo 2015, summer festival about Japanese culture

- 250k tickets over 4 days
- €130 per person
- More than €8M every year

Mangaki

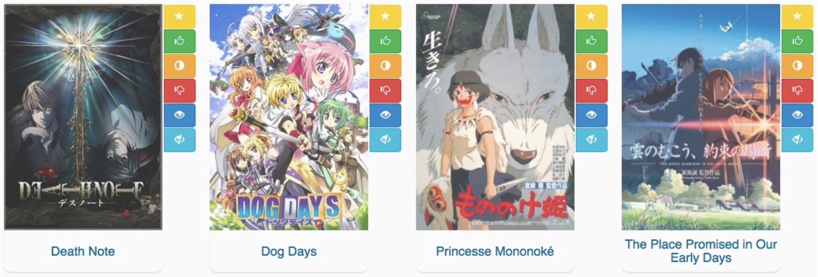
Let's use algorithms to discover pearls of Japanese culture!

23 members:
3 mad developers
1 graphic designer
1 intern

- Feb 2014 - PhD about Adaptive Testing
- Oct 2014 - Mangaki
- Oct 2015 - Student Demo Cup winner, Microsoft prize
- Feb 2016 - Japanese Culture Embassy Prize, Paris

We're flying to Tokyo!

Mangaki



- 2000 users
- 150 $\rightarrow$ 14000 works *anime / manga / OST*
- 290000 ratings *fav / like / dislike / neutral / willsee / wontsee*
- People rate a few works *Preference Elicitation*
- And receive recommendations *Collaborative Filtering*

Collaborative Filtering

Problem

- Users $u = 1, \ldots, n$ and items $i = 1, \ldots, m$
- Every user u rates some items $\mathcal{R}_u$
  (r_{ui}: rating of user u on item i)
- How to guess unknown ratings?

k-nearest neighbor

- Similarity score between users:

$$score(u, v) = \frac{\mathcal{R}_u \cdot \mathcal{R}_v}{||\mathcal{R}_u|| \cdot ||\mathcal{R}_v||}.$$

- Let us find the k nearest neighbors of the user
- And recommend what they liked that he did not rate

Preference Elicitation

Problem

What questions to ask adaptively to a new user?

4 decks

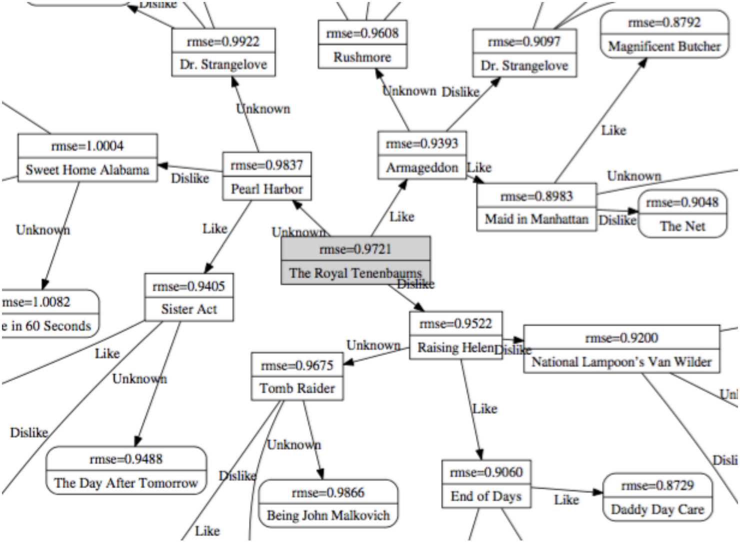
- Popularity
- Controversy (Reddit)

$$controversy(L, D) = (L + D)^{min(L/D, D/L)}$$

- Most liked
- *Precious pearls*: few rates but almost no dislike

Problem: most people cannot rate the controversial items

Yahoo's Bootstrapping Decision Trees



Good balance between like, dislike and unknown outcomes.

Matrix Completion

If the matrix of ratings is assumed low rank:

$$M = \begin{pmatrix} \mathcal{R}_1 \\ \mathcal{R}_2 \\ \vdots \\ \mathcal{R}_n \end{pmatrix} = \boxed{} = \boxed{C}\,\boxed{P}$$

Every line $\mathcal{R}_u$ is a linear combination of few profiles P.

1. Explicable profiles

If P	P_1 : adventure	P_2 : romance	P_3 : plot twist
And C_u	0,2	$-0{,}5$	0,6

$\Rightarrow$ u likes a bit adventure, dislikes romance, loves plot twists.

2. We get user2vec & item2vec in the same space!

$\Rightarrow$ An user likes items that are close to him.

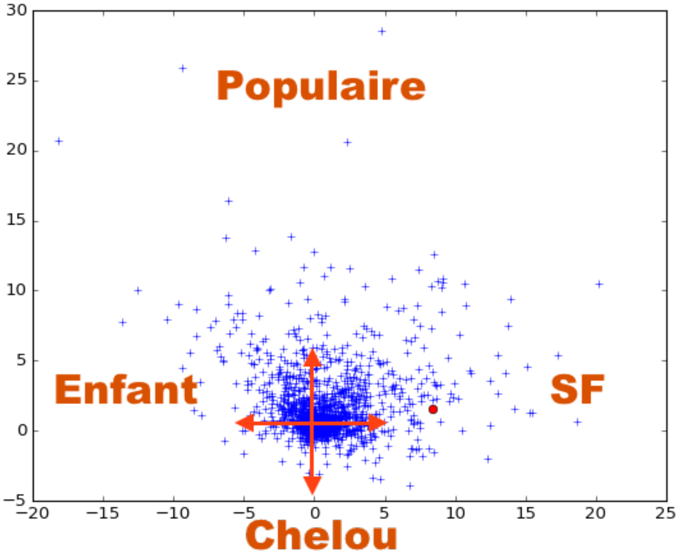
Mangaki's Explained Profiles

P_1 loves stories for adults (Ghibli/SF), hates stories for teens

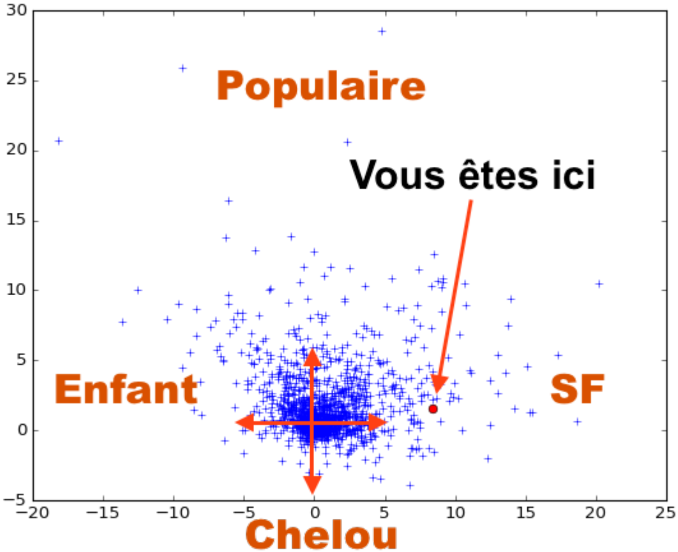
Paprika : *In the near future, a machine that allows to enter another person's dreams for psychotherapy treatment has been robbed. The doctors investigate.*

P_2 likes the most popular works, hates really weird works

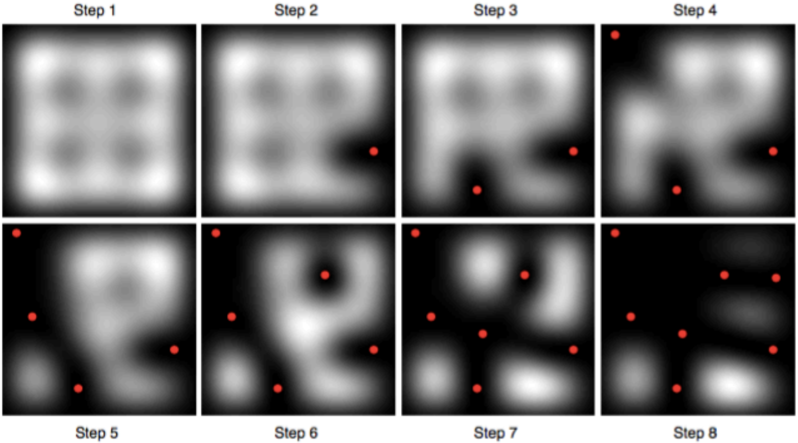
Same-family homosexual romances:
- Mira is a high school student in love with his father, Kyōsuke.
- They are involved both in a romantic and sexual relationship.
- Trouble arises when Mira finds adoption papers.
- Mira thinks [his dad] is cheating on him with a girl, which turns out to be his mother.
- Meanwhile, Mira is chased by his best friend, who is also in love with him.
- And the end, it turns out that Kyōsuke is his uncle and everything is right again.


30
25
20
15
10
5
0
−5
−20
−15
−10
−5
0
5
10
15
20
25
Populaire
Enfant
SF
Chelou

user2vec: plotting every user on a map



Let's sample a few items that are far from each other and ask them to the user.



Determinantal Point Processes

We want to sample over n items
$K : n \times n$ similarity matrix over items (positive semidefinite)

P is a determinantal point process if Y is drawn such that:

$$\forall A \subset Y, \quad P(A \subseteq Y) \propto det(K_A)$$

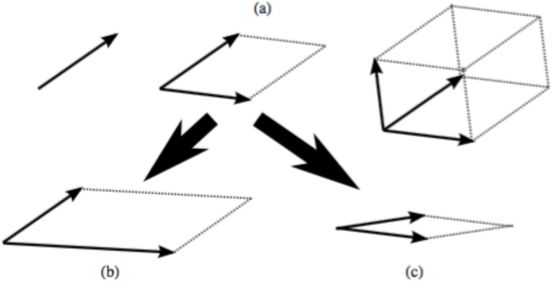
Example

$$K = \begin{pmatrix} 1 & 2 & 3 & 4 \\ 2 & 5 & 6 & 7 \\ 3 & 6 & 8 & 9 \\ 4 & 7 & 9 & 0 \end{pmatrix}$$

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

$$K_A = det \begin{pmatrix} 1 & 2 & 4 \\ 2 & 5 & 7 \\ 4 & 7 & 0 \end{pmatrix}$$

Link with diversity



- The determinant is the volume of the vectors
- Non-correlated (diverse) vectors will increase the volume
- We need to sample k diverse elements efficiently

Results

We assume the eigenvalues of K are known.

Kulesza & Taskar, ICML 2011

Sampling k diverse elements from a DPP of size n has complexity $O(nk^2)$.

Kang (Samsung), NIPS 2013

We found an algorithm ϵ-close with complexity $O(k^3 \log(k/\epsilon))$.

Rebeschini & Karbasi, COLT 2015

No, you did not.
Your proof of complexity is false, but at least your algorithm samples correctly.

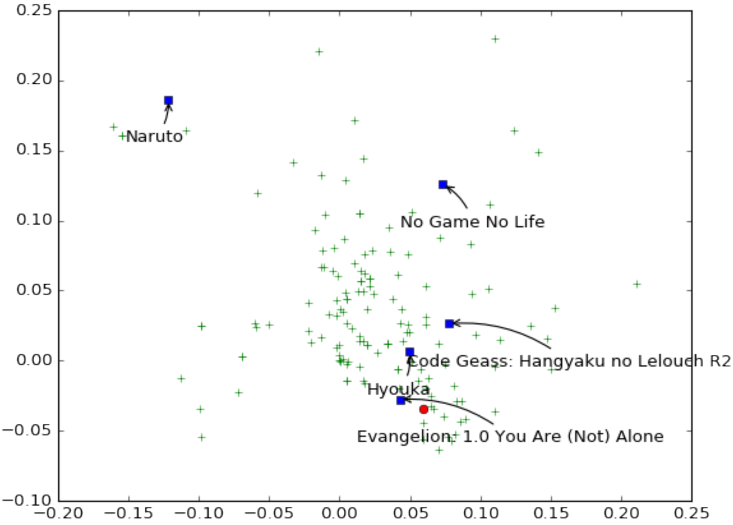
Vie, RecSysFR 2016

Please calm down guys, we're all friends here.

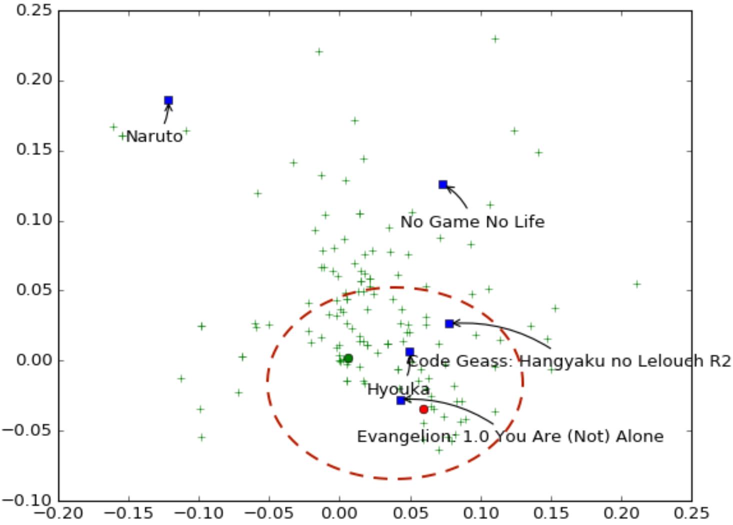
Work in progress: Preference Elicitation in Mangaki

- Keep the 100 most popular works
- Pick 5 diverse works using k-DPP
- Ask the user to rate them
- Estimate the user's vector
- Filter less informative works accordingly
- Repeat.

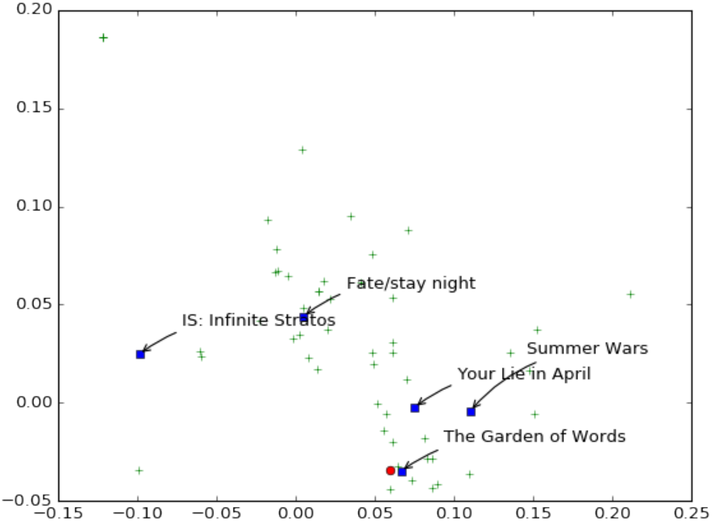
Preference elicitation: sampling a diverse subset



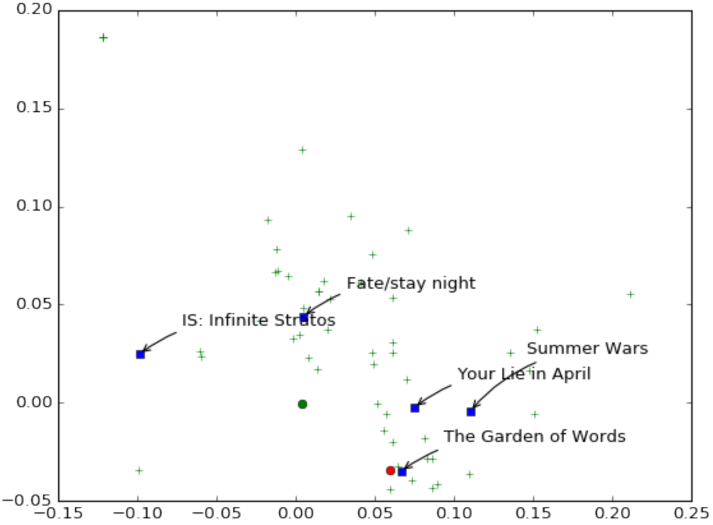
Estimating according to the answers and filtering



Sampling again



Refining estimate



@MangakiFR

research.mangaki.fr

We're on GitHub!

MANGAKI

LA RECOMMANDATION FRANCOPHONE D'ANIME ET MANGA