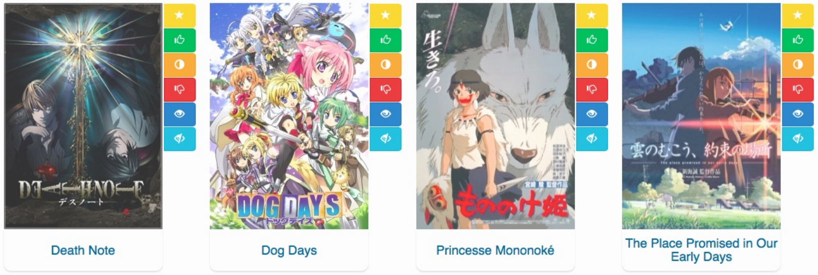
# Using Ratings & Posters for Anime & Manga Recommendations

Jill-Jênn Vie[1,3] Florian Yger[2] Ryan Lahfa[3] Basile Clement[3]
Kévin Cocchi[3] Thomas Chalumeau[3] Hisashi Kashima[1,4]

[1] RIKEN Center for Advanced Intelligence Project (Tokyo)
[2] Université Paris-Dauphine (France)
[3] Mangaki (Paris, France)
[4] Kyoto University

# Mangaki.fr

- User can rate anime or manga (works)
- And receive recommendations
- And reorder their watchlist




- Code is 100% on GitHub
- Awards from Microsoft and Japan Foundation
- Organized a data challenge with Kyoto University

# RIKEN Center for Advanced Intelligence Project


- New AI lab near Tokyo Station (opened in 2016)
- 8 accepted papers at NIPS 2017

# Authors

Jill-Jênn Vie Florian Yger Ryan Lahfa Hisashi Kashima

- Florian Yger was visiting RIKEN AIP
- Kévin Cocchi & Thomas Chalumeau were interns at Mangaki

# Outline

## 1. Usual algorithms for recommender systems

- Content-based
- Collaborative filtering

## 2. Our method

- Extracting tags from posters
- Blending models

## 3. Experiments

- Dataset: Mangaki
- Results

# Recommender Systems

## Problem

- Every user rates few items (1 %)
- How to infer missing ratings?

## Example

| |  |  |  |  |
|---|---|---|---|---|
| Sacha | ? | 5 | 2 | ? |
| Ondine | 4 | 1 | ? | 5 |
| Pierre | 3 | 3 | 1 | 4 |
| Joëlle | 5 | ? | 2 | ? |

# Recommender Systems

## Problem

- Every user rates few items (1 %)
- How to infer missing ratings?

## Example





| | | | | |
|---|---|---|---|---|
| Sacha | 3 | 5 | 2 | 2 |
| Ondine | 4 | 1 | 4 | 5 |
| Pierre | 3 | 3 | 1 | 4 |
| Joëlle | 5 | 2 | 2 | 5 |

## Usual techniques

Content-based (work features: directors, genre, etc.)

- Linear regression
- Sparse linear regression (LASSO)

Collaborative filtering (solely based on ratings)

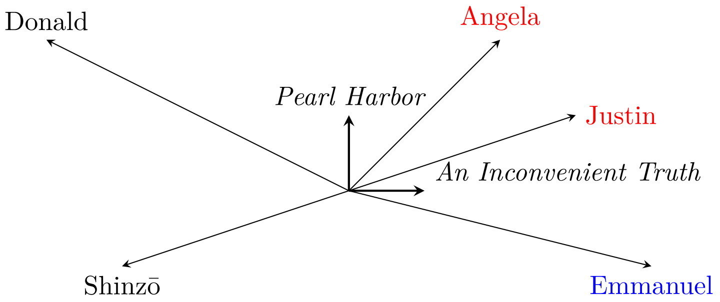
- $K$-nearest neighbors
- Matrix factorization:
  - Singular value decomposition
  - Alternating least squares
  - Stochastic gradient descent

Hybrid recommender systems (combine those two)

- The proposed method

# Example: $K$-Nearest Neighbors

| **Ratings** | *Paprika* | *Pearl Harbor* | *An Inconvenient Truth* |
|---|---|---|---|
| Justin | 3 | 1 | 3 |
| Angela | ? | 2 | 2 |
| Donald | $-3$ | 2 | $-4$ |
| Emmanuel | ? | $-1$ | 4 |
| Shinzō | 4 | $-1$ | $-3$ |



# Example: $K$-Nearest Neighbors

| **Ratings** | *Paprika* | *Pearl Harbor* | *An Inconvenient Truth* |
|---|---|---|---|
| Justin | 3 | 1 | 3 |
| Angela | ? | 2 | 2 |
| Donald | $-3$ | 2 | $-4$ |
| Emmanuel | 3,5 | $-1$ | 4 |
| Shinzō | 4 | $-1$ | $-3$ |

| **Similarity** | Justin | Angela | Donald | Emmanuel | Shinzo |
|---|---|---|---|---|---|
| Justin | 1 | 0,649 | $-0{,}809$ | 0,612 | 0,090 |
| Angela | 0,649 | 1 | $-0{,}263$ | 0,514 | $-0{,}555$ |
| Donald | $-0{,}809$ | $-0{,}263$ | 1 | $-0{,}811$ | $-0{,}073$ |
| Emmanuel | 0,612 | 0,514 | $-0{,}811$ | 1 | $-0{,}523$ |
| Shinzō | 0,090 | $-0{,}555$ | $-0{,}073$ | $-0{,}523$ | 1 |

# Matrix factorization $\rightarrow$ reduce dimension to generalize

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

$R$: 2k users $\times$ 15k works $\iff \begin{cases} C\text{: 2k users} \times \text{20 profiles} \\ P\text{: 20 profiles} \times \text{15k works} \end{cases}$

$\mathcal{R}_{\text{Bob}}$ is a linear combination of profiles $P_1$, $P_2$, etc..

## Interpreting Key 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, hates romance, loves plot twists.

# Weighted Alternating Least Squares (Zhou, 2008)

$R$ ratings, $U$ user features, $V$ work features.

$$R = UV^T \qquad \Rightarrow \qquad r_{ij} \simeq \hat{r}_{ij}^{ALS} \triangleq U_i \cdot V_j.$$

## Objective function to minimize

$$U, V \mapsto \sum_{i,j \text{ known}} (r_{ij} - U_i \cdot V_j)^2 + \lambda \left( \sum_i N_i ||U_i||^2 + \sum_j M_j ||V_j||^2 \right)$$

where:

- $N_i$: number of ratings by user $i$
- $M_j$: number of ratings for item $j$

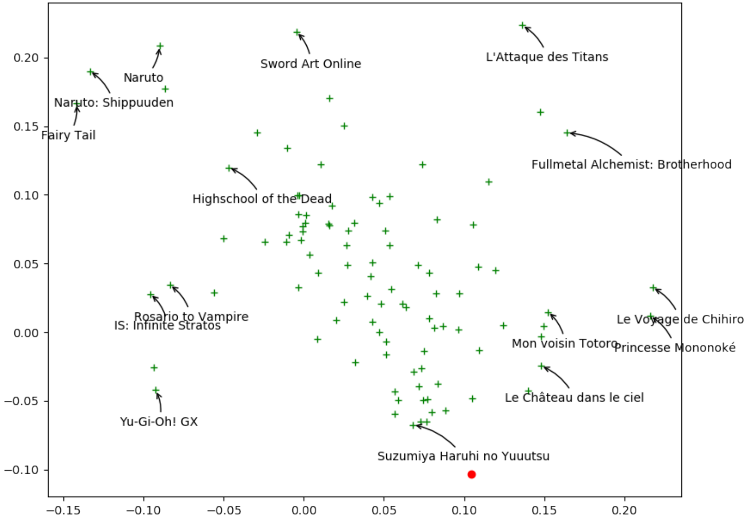
## Algorithm

Until convergence ($\sim$ 10 iterations):

- Fix $U$ find $V$ (just linear regression $\rightarrow$ least squares)
- Fix $V$ find $U$

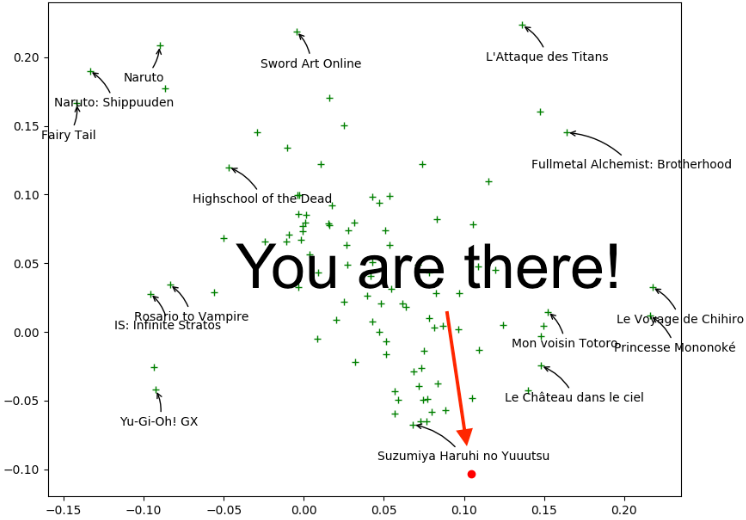
# Visualizing first two components of anime $V_j$

Closer points mean similar taste



# Find your taste by plotting first two columns of $U_i$

You will like anime that are in your direction



# Drawback with collaborative filtering

## Issue: Item Cold-Start

- If no ratings are available for a work $j$
  $\Rightarrow$ Its features $V_j$ cannot be trained :-(

No way to distinguish between unrated works.

## But we have posters!

- On Mangaki, almost all works have a poster
- How to extract information?

# Illustration2Vec (Saito and Matsui, 2015)


Prediction results

| # | General Tag | Confidence |
|---|---|---|
| 1. | 1girl | 86.1% |
| 2. | thighhighs | 84.0% |
| 3. | solo | 79.2% |
| 4. | red hair | 73.1% |
| 5. | long hair | 66.4% |
| 6. | breasts | 53.7% |
| 7. | gloves | 38.0% |
| 8. | weapon | 34.0% |
| 9. | elbow gloves | 28.3% |
| 10. | high heels | 14.0% |
| 11. | tattoo | 10.9% |
| # | **Character Tag** | |
| # | **Copyright Tag** | |
| # | **Rating** | |
| 1. | safe | 68.4% |
| 2. | questionable | 29.3% |
| 3. | explicit | 1.92% |

- CNN (VGG-16) pretrained on ImageNet, trained on Danbooru (1.5M illustrations with tags)
- 502 most frequent tags kept, outputs tag weights

# LASSO for sparse linear regression

$T$ matrix of 15000 works $\times$ 502 tags ($t_{jk}$: tag $k$ appears in item $j$)

- Each user is described by its preferences $P_i$
  $\rightarrow$ a sparse row of weights over tags.
- Estimate user preferences $P_i$ such that

$$r_{ij} \simeq \hat{r}_{ij}^{LASSO} \triangleq P_i T_j^T.$$

## Least Absolute Shrinkage and Selection Operator (LASSO)

$$P_i \mapsto \frac{1}{2N_i} \|\mathcal{R}_i - P_i T^T\|_2^2 + \alpha \|P_i\|_1.$$

where $N_i$ is the number of items rated by user $i$.

## Interpretation and explanation of user preferences

- *You seem to like magical girls but not blonde hair*
  $\Rightarrow$ *Look! All of them are brown hair! Buy now!*

# Combine models

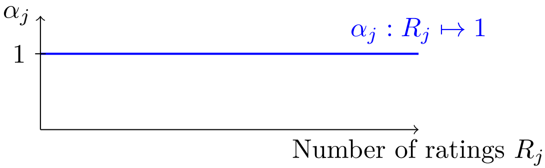
Which model should be choose between ALS and LASSO?

**Answer** Both!

**Methods** boosting, bagging, model stacking, blending.

**Idea** find $\alpha_j$ s.t. $\hat{r}_{ij} \triangleq \alpha_j \hat{r}_{ij}^{ALS} + (1 - \alpha_j)\hat{r}_{ij}^{LASSO}$.

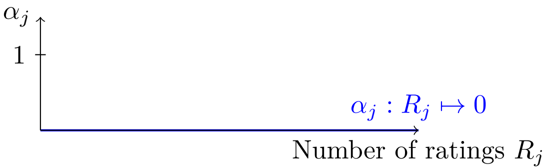
# Examples of $\alpha_j$




Mimics ALS

$$\hat{r}_{ij} \triangleq 1\hat{r}_{ij}^{ALS} + 0\hat{r}_{ij}^{LASSO}.$$

# Examples of $\alpha_j$

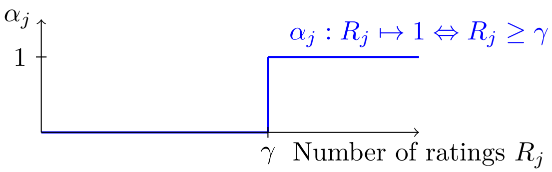



Mimics LASSO

$$\hat{r}_{ij} \triangleq 0\hat{r}_{ij}^{ALS} + 1\hat{r}_{ij}^{LASSO}.$$

We call this gate the Steins;Gate.

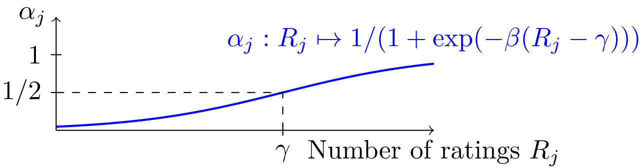
# Examples of $\alpha_j$




$$\hat{r}_{ij}^{BALSE} = \begin{cases} \hat{r}_{ij}^{ALS} & \text{if item } j \text{ was rated at least } \gamma \text{ times} \\ \hat{r}_{ij}^{LASSO} & \text{otherwise} \end{cases}$$

But we can't: Not differentiable!

# Examples of $\alpha_j$

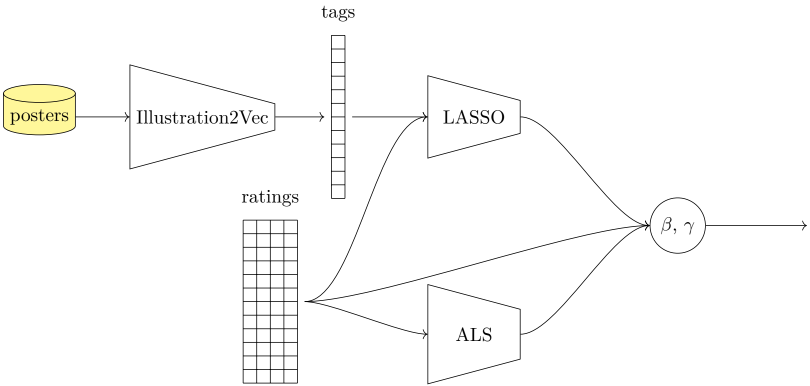



$$\hat{r}_{ij}^{BALSE} = \sigma(\beta(R_j-\gamma))\hat{r}_{ij}^{ALS} + (1-\sigma(\beta(R_j-\gamma)))\,\hat{r}_{ij}^{LASSO}$$

$\beta$ and $\gamma$ are learned by stochastic gradient descent.

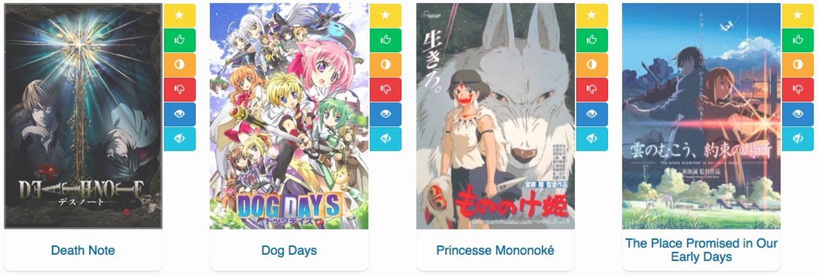
We call this gate the Steins;Gate.

# Blended Alternate Least Squares with Explanation




We call this model BALSE.

# Dataset: Mangaki




- 2300 users
- 15000 works — *anime / manga / OST*
- 340000 ratings — *fav / like / dislike / neutral / willsee / wontsee*

# Evaluation: Root Mean Squared Error (RMSE)

If we predict $\hat{r}_{ij}$ for each user-work pair $(i, j)$ to test among $n$, while truth is $r_{ij}$:

$$RMSE(\hat{r}, r) = \sqrt{\frac{1}{n} \sum_{i,j} (\hat{r}_{ij} - r_{ij})^2}.$$

# Cross-validation

- 80% of the ratings are used for training
- 20% of the ratings are kept for testing

Differents sets of items:

- Whole test set of works
- 1000 works least rated (1.5%)
- Cold-start: works not seen in the training set (only the posters)

# Results

| RMSE | Test set | 1000 least rated (1.5%) | Cold-start items |
|---|---|---|---|
| ALS | 1.157 | 1.299 | 1.493 |
| LASSO | 1.446 | 1.347 | 1.358 |
| BALSE | **1.150** | **1.247** | **1.316** |

# Summing up

We presented BALSE, a model that:

- uses information in the ratings (collaborative filtering)
- uses information in the posters using CNNs (content-based)
- combine them in a nonlinear way

to improve the recommendations, and explain them.

## Further work

- Use latent features (not only tags) of the posters (IJCAI 2016)
- End-to-end training (not separately)

# Thank you!

MANGAKI
LA RECOMMANDATION FRANCOPHONE D'ANIME ET MANGA

Try it: https://mangaki.fr

Twitter: @MangakiFR

## Read the article

Using Posters to Recommend Anime and Mangas in a Cold-Start Scenario
github.com/mangaki/balse (PDF on arXiv, front page of HNews)

## Results of Mangaki Data Challenge: research.mangaki.fr

1. Ronnie Wang (Microsoft Suzhou, China)
2. Kento Nozawa (Tsukuba University, Japan)
3. Jo Takano (Kobe University, Japan)