

Fouille de Données et Media Sociaux  
Cours 3  
Recommendation  
Master DAC Data Science  
UPMC - LIP6

Ludovic Denoyer

4 octobre 2015

# La distribution

La mise à disposition...

Où et quand ?

..d'un produit ou service...

Quoi ?

..à un intermédiaire ou consommateur final...

Qui ?

# La distribution

Distribution "classique"

Visite du point de vente

Distribution "numérique" (en ligne)

Produits :

- physiques
- numériques

dans un catalogue de milliers ou millions de références.

# Recommandation

## Séries et TV

Parce que cela vous intéressera :



## Tendances actuelles



## Animation

Parce que cela vous intéressera :



# Recommandation

[Feuilleter ↴](#)

**PATTERN RECOGNITION AND MACHINE LEARNING** CHRISTOPHER M. BISHOP

**Pattern Recognition And Machine Learning** (Anglais) Relié – 17 août 2006  
de Christopher M. Bishop \* (Auteur)  
 4 commentaires client

[Voir les formats et éditions](#)

Relié EUR 60,42	Relié à partir de EUR 179,81	Broché à partir de EUR 22,89
7 d'occasion à partir de EUR 52,11 16 neufs à partir de EUR 52,11	1 d'occasion à partir de EUR 179,81 2 neufs à partir de EUR 219,81	1 d'occasion à partir de EUR 27,53 15 neufs à partir de EUR 22,89

Voulez-vous le faire livrer le lundi 5 oct.? Commandez-le dans les 22 h et 34 mins et choisissez la Livraison en soirée au cours de votre commande. [En savoir plus.](#)

Livres anglais et étrangers  
Lisez en version originale. [Cliquez ici](#)



[Voir les 3 images](#)

## Offres spéciales et liens associés

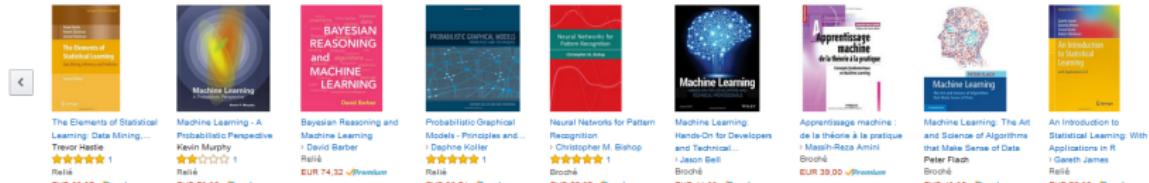
- [Rentrée Scolaire 2015 : découvrez toutes nos sélections et promotions dans notre boutique Rentrée 2015.](#)
- [Recevez un code promotionnel de 10€ pour l'achat d'un chèque-cadeau Amazon.fr d'une valeur de 50€.](#)

## Produits fréquemment achetés ensemble



- [Cet article : Pattern Recognition And Machine Learning de Christopher M. Bishop Relié EUR 60,42](#)
- [Machine Learning - A Probabilistic Perspective de Kevin Murphy Relié EUR 70,32](#)
- [The Elements of Statistical Learning: Data Mining, Inference, and Prediction de Trevor Hastie Relié EUR 69,37](#)

## Les clients ayant acheté cet article ont également acheté



The Elements of Statistical Learning Data Mining, Inference, and Prediction Trevor Hastie ★★★★ 1 Relié EUR 69,37 ✓ Disponible

Machine Learning - A Probabilistic Perspective Kevin Murphy ★★★★★ 1 Relié EUR 70,32 ✓ Disponible

Bayesian Reasoning and Machine Learning David Barber Relié EUR 74,32 ✓ Premium

Probabilistic Graphical Models - Principles and Techniques David M. J. Beal Relié EUR 69,74 ✓ Disponible

Neural Networks for Pattern Recognition Christopher M. Bishop Relié EUR 59,11 ✓ Disponible

Machine Learning: Hands-On for Developers and Technical Professionals Jason Bell Broché EUR 45,45 ✓ Disponible

Apprentissage machine de la théorie à la pratique Massih-Reza Amini Broché EUR 39,00 ✓ Premium

Machine Learning: The Art and Science of Algorithms That Make Sense of Data Peter Flach Broché EUR 45,05 ✓ Disponible

An Introduction to Statistical Learning Gareth James Relié EUR 74,00 ✓ Disponible

Ludovic Denoyer

FDMS

4 octobre 2015

5 / 45

# Recommandation

vache folle

L'abat était interdit à la vente depuis la vache folle

Iépolinaire.fr - 9 sept. 2015  
Jean-Pierre Ribière constate que l'intestin grêle de veau retrouve vite son public?! - Photo Pascal Lachenaud ...

Leatitia et Marie-Ange Casta : Duo irrésistible pour les 95 ans de ...  
Pure People - il y a 3 heures  
Le vert ardéole, les jaunes miel et moutarde, et un imprimé peau de vache... pour la folle soirée des 95 ans de l'édition française de Vogue.

L'Europe subit les conséquences de l'affaire Volkswagen

State.fr - 1 oct. 2015  
Brookings Institution souligne la comparaison avec l'affaire de la vache folle, où des règles communes avaient été trouvées mais où les lobbies ...

Irlande: Premier cas de vache folle détecté depuis 2013

20minutes.fr - 25 juin 2015  
Le premier cas depuis 2013 d'encéphalopathie spongiforme bovine (ESB), communément appelé vache folle, a été détecté en Irlande, ont ...

Premier cas de vache folle détecté en Irlande depuis 2013

Exhaustif - Le Figaro - 26 juin 2015  
Explorer en détail (37 autres articles)

Maladie de la vache folle : enfin un espoir de traitement ?

Sciences et Avenir - 6 août 2015  
... protéines considérées comme des agents de l'encéphalopathie spongiforme bovine (ESB) ou maladie de la vache folle, ouvrant la voie à un ...

Vache folle : un espoir de traitement

Cité à de nombreuses reprises - Pourquoi Docteur ? - 6 août 2015  
Explorer en détail (15 autres articles)

# Recommandation

ALERTE INFO - 16 morts dans des crues dans le sud-est de la France



## Ligue 1 : suivez Monaco - Rennes en direct

Les Bretons, fortement diminués, défient une équipe monégasque qui espère gagner pour recoller à la tête du classement. [Le live commenté](#) »

1-5 sur 25



De nombreuses victimes



Monaco - Rennes en direct



Un tapis rouge inoubliable



Une innovation très pratique



Elle ne s'attendait pas à ça



# Recommandation

- Recommandation de films (Netflix)
- Recommandation de produits (Amazon)
- Recommandation de pages Web (Google)
- Recommandation sociale (Facebook, Linkedin)
- Recommandation de news (Yahoo !)
- Recommandation de rencontres (OK Cupid)
- Recommandation de publicité (Google, Yahoo !, ...)
- Recommandation de musique (Deezer, ...)
- ...

# Recommandation

## Vue formelle

Problème qui contient :

- Un utilisateur
- Un ensemble d'objets (ou items)
- Un contexte de recherche
- Une interface homme-machine

# Enjeux

## Enjeux

AMAZON augmente ses ventes de 40 % lorsque le moteur de recommandation est utilisé sur la vente des produits à faible rotation et faible visibilité

Et d'autres....

- 1/3 des startups "prometteuses" à Paris sont directement ou indirectement basées sur des problématiques de recommandation

## Exemple : le challenge Netflix

(Wikipedia)

The Netflix Prize was an open competition for the best collaborative filtering algorithm to predict user ratings for films, based on previous ratings without any other information about the users or films, i.e. without the users or the films being identified except by numbers assigned for the contest.

The competition was held by Netflix, an online DVD-rental and online video streaming service, and was open to anyone not connected with Netflix (current and former employees, agents, close relatives of Netflix employees, etc.) or a resident of Cuba, Iran, Syria, North Korea, Burma or Sudan.<sup>[1]</sup> On 21 September 2009, the grand prize of **US\$1,000,000** was given to the BellKor's Pragmatic Chaos team which bested Netflix's own algorithm for predicting ratings by 10.06%

## Exemple : le challenge Netflix

Training data

user	movie	date	score
1	21	5/7/02	1
1	213	8/2/04	5
2	345	3/6/01	4
2	123	5/1/05	4
2	768	7/15/02	3
3	76	1/22/01	5
4	45	8/3/00	4
5	568	9/10/05	1
5	342	3/5/03	2
5	234	12/28/00	2
6	76	8/11/02	5
6	56	6/15/03	4

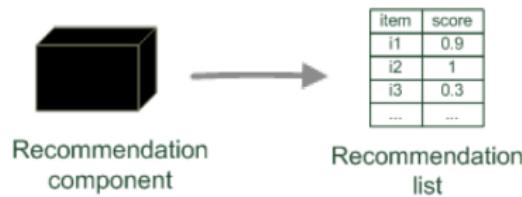
Test data

user	movie	date	score
1	62	1/6/05	?
1	96	9/13/04	?
2	7	8/18/05	?
2	3	11/22/05	?
3	47	6/13/02	?
3	15	8/12/01	?
4	41	9/1/00	?
4	28	8/27/05	?
5	93	4/4/05	?
5	74	7/16/03	?
6	69	2/14/04	?
6	83	10/3/03	?

480,189 User ID's, 17,770 movies, 100,480,507 ratings collected from October 1998 to December 2005.

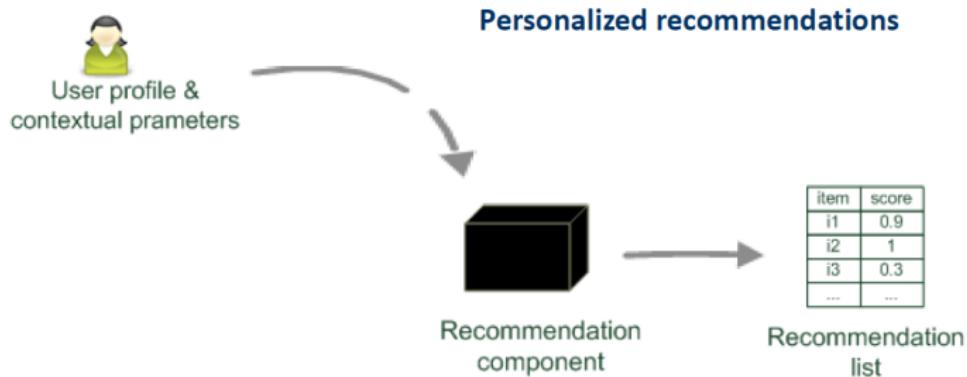
# Paradigmes de Recommandation

**Recommender systems reduce  
information overload by estimating  
relevance**



Source : Tutorial Recommender Systems, Dietmar Jannach, Gerhard Friedrich

# Paradigmes de Recommandation



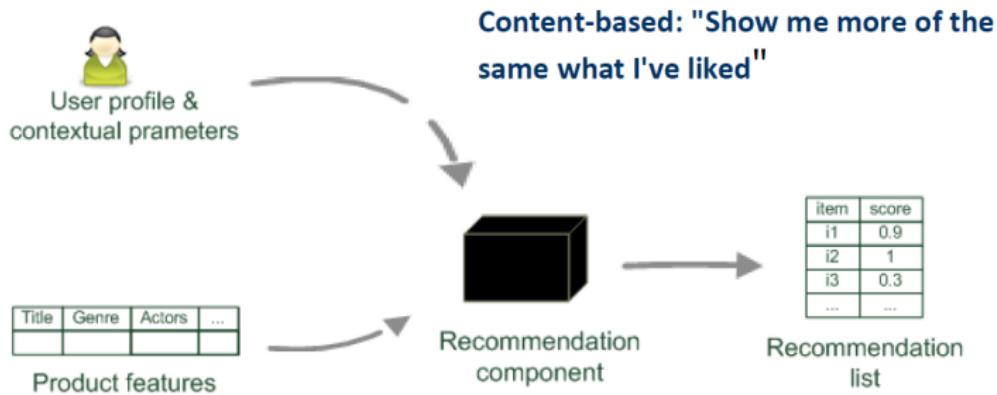
Source : Tutorial Recommender Systems, Dietmar Jannach, Gerhard Friedrich

# Paradigmes de Recommandation



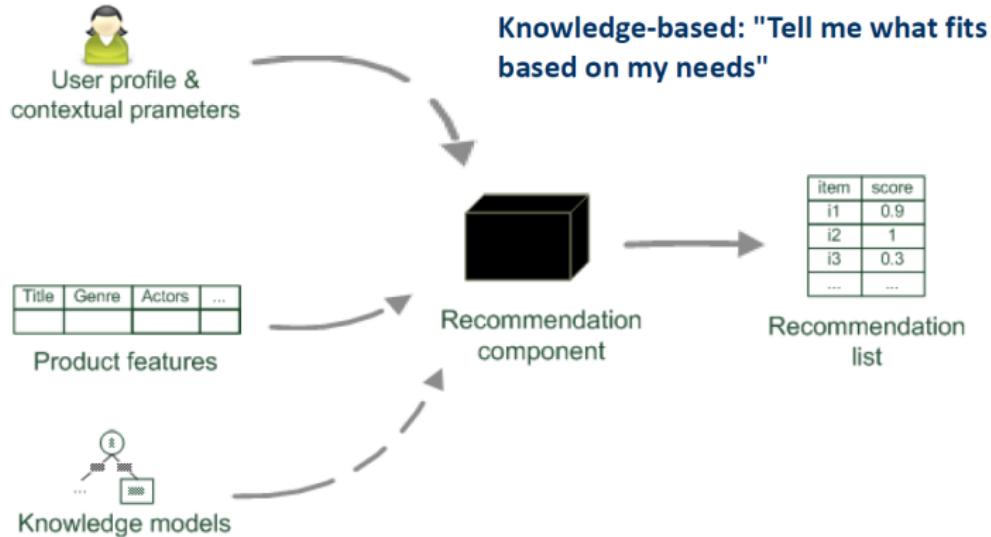
Source : Tutorial Recommender Systems, Dietmar Jannach, Gerhard Friedrich

# Paradigmes de Recommandation



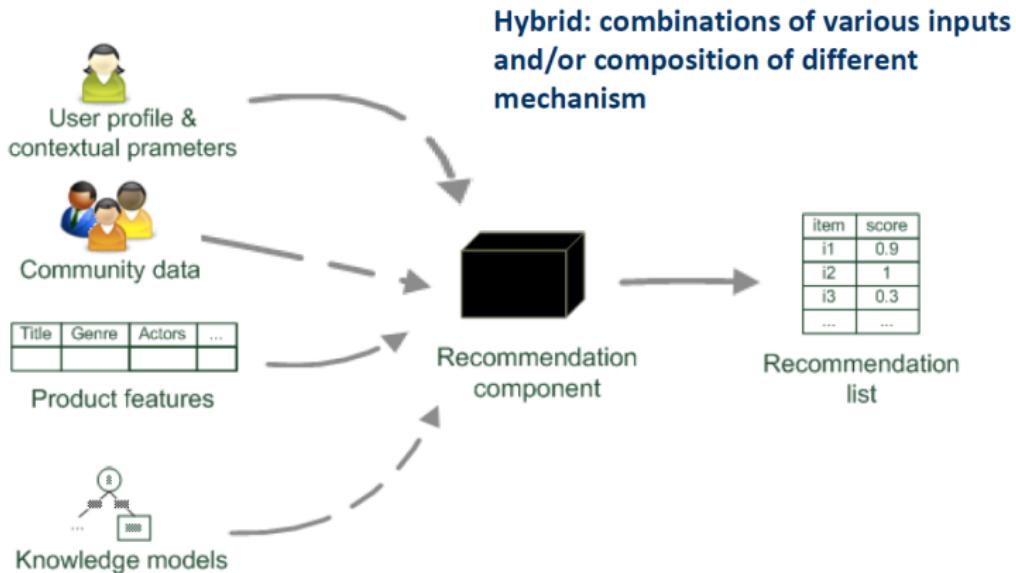
Source : Tutorial Recommender Systems, Dietmar Jannach, Gerhard Friedrich

# Paradigmes de Recommandation



Source : Tutorial Recommender Systems, Dietmar Jannach, Gerhard Friedrich

# Paradigmes de Recommandation



Source : Tutorial Recommender Systems, Dietmar Jannach, Gerhard Friedrich

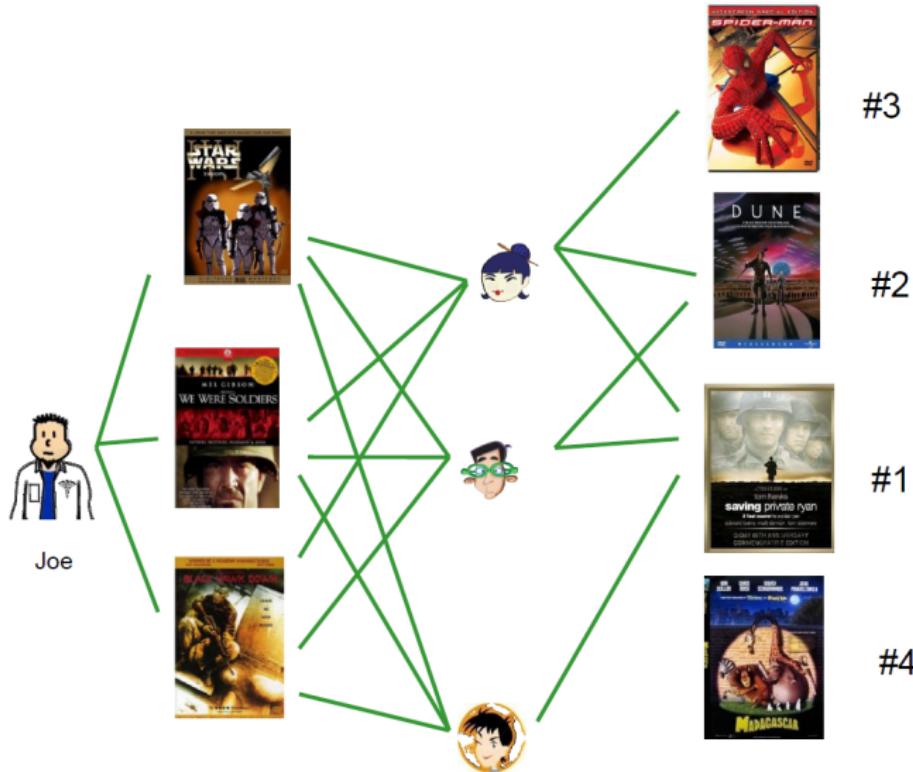
# Filtrage Collaboratif

## Définition

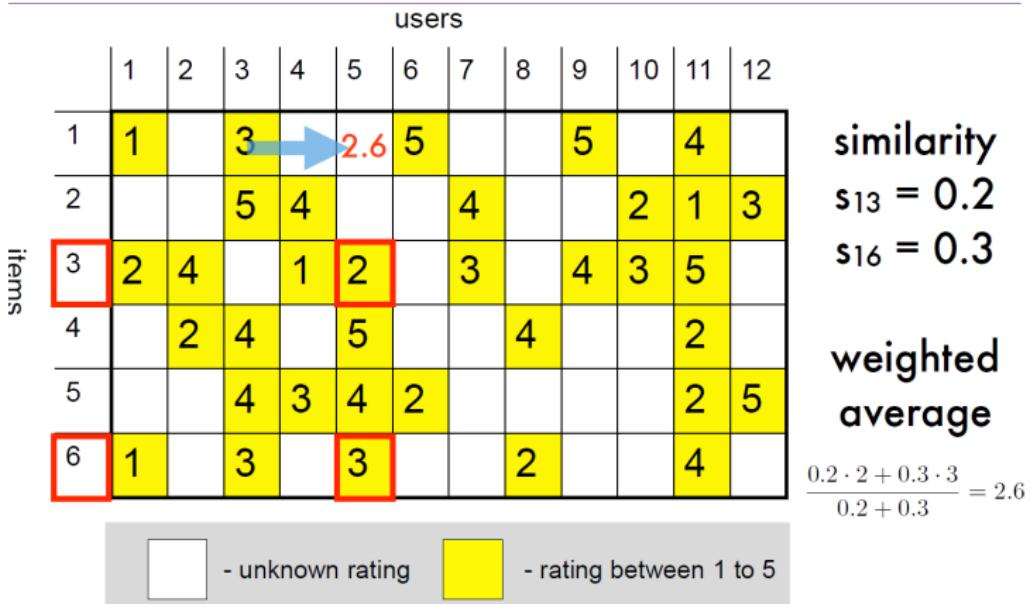
Collaborative filtering is a method of making automatic predictions (filtering) about the interests of a user by collecting preferences or taste information from many users (collaborating). The underlying assumption of the collaborative filtering approach is that if a person A has the same opinion as a person B on an issue, A is more likely to have B's opinion on a different issue x than to have the opinion on x of a person chosen randomly.

Source : Wikipedia

# Similarité d'utilisateurs



# Similarité d'utilisateurs



Source : Scalable Machine Learning - Alex Smola - Yahoo !

# Similarité d'utilisateurs

## Avantages

- Facile
- Prise en compte de nouveaux utilisateurs et items
- Facile à expliquer

# Corrélation de Pearson

Comment mesurer la similarité des utilisateurs ?

$$Sim(a, b) = \frac{\sum_p (r_{a,p} - \bar{r}_a)(r_{b,p} - \bar{r}_b)}{\sqrt{\sum_p (r_{a,p} - \bar{r}_a)^2} \sqrt{\sum_p (r_{b,p} - \bar{r}_b)^2}} \quad (1)$$

- $p \in \mathcal{P}$  l'ensemble des items
- $r_{a,p}$  : la note de l'utilisateur  $a$  sur l'item  $p$
- $\bar{r}_a$  : la note moyenne de l'utilisateur  $a$

Prédiction

$$pred(a, p) = \bar{r}_a + \frac{\sum_b sim(a, b) * (r_{b,p} - \bar{r}_b)}{\sum_b sim(a, b)} \quad (2)$$

# Similarité d'items

Comment mesurer la similarité des items ?

$$Sim(i, j) = \frac{\sum_u (r_{u,i} - \bar{r}_i)(r_{u,j} - \bar{r}_j)}{\sqrt{\sum_p (r_{u,i} - \bar{r}_i)^2} \sqrt{\sum_p (r_{u,j} - \bar{r}_j)^2}} \quad (3)$$

- $u \in \mathcal{U}$  l'ensemble des utilisateurs
- $r_{u,i}$  : la note de l'utilisateur  $u$  sur l'item  $i$
- $\bar{r}_i$  : la note moyenne de l'item  $i$

Prédiction

$$pred(u, i) = \bar{r}_i + \frac{\sum_j sim(i, j) * (r_{u,j} - \bar{r}_j)}{\sum_b sim(i, j)} \quad (4)$$

# Plongement dans des espaces de faible dimension

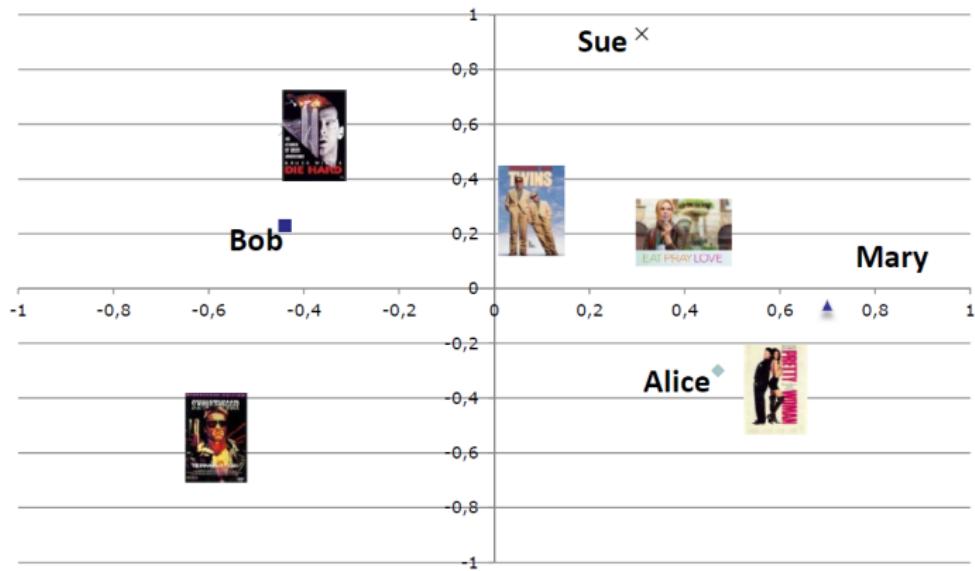
## Inconvénient des méthodes précédentes

- Le calcul des similarités entre tous les utilisateurs (en prédition) est un calcul très long
- Le calcul des similarités entre tous les items (en prédition) est un calcul très long

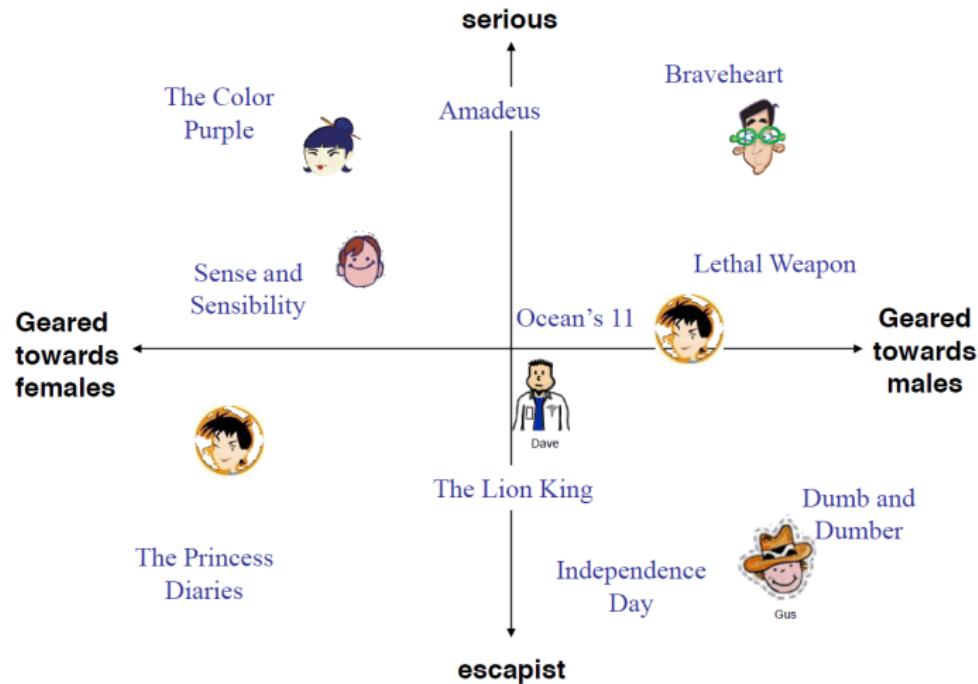
## Solutions

- Utilisation des  $k$ -plus proches voisins - calcul a priori
- Plongement des utilisateurs et des items dans un espace de faible dimension (embedding)

# Low Dimensional Embedding of the users



# Low Dimensional Embedding of the users



# Low Dimensional Embedding of the users

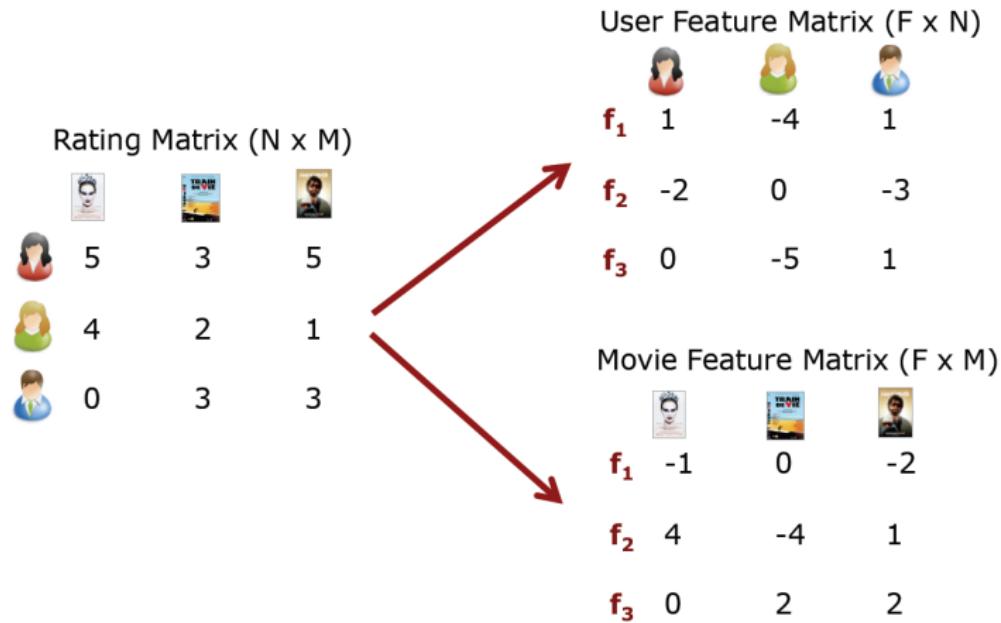
$U_k$	Dim1	Dim2
Alice	0.47	-0.30
Bob	-0.44	0.23
Mary	0.70	-0.06

$V_k^T$					
Dim1	-0.44	-0.57	0.06	0.38	0.57
Dim2	0.58	-0.66	0.26	0.18	-0.36

Le score final est calculé à l'aide d'un similarité entre utilisateurs et items :

$$r_{u,i} = \langle p_u; q_i \rangle \quad (5)$$

# Factorisation de matrices



# Factorisation de matrices

$$\text{User Feature Matrix (F x N)} \quad \begin{matrix} f_1 & 1 & -4 & 1 \\ f_2 & -2 & 0 & -3 \\ f_3 & 0 & -5 & 1 \end{matrix}^T \cdot \text{Movie Feature Matrix (F x M)} \quad \begin{matrix} f_1 & -1 & 0 & -2 \\ f_2 & 4 & -4 & 1 \\ f_3 & 0 & 2 & 2 \end{matrix} = \text{Rating Matrix (N x M)} \quad \begin{matrix} \text{User 1} & \text{Movie 1} & \text{Movie 2} & \text{Movie 3} \\ \text{User 1} & 5 & 3 & 5 \\ \text{User 2} & 4 & 2 & 1 \\ \text{User 3} & 0 & 3 & 3 \end{matrix}$$

# Factorisation de matrices

## Problème général

$$\text{Minimise } C(P, Q) = \|X - PQ\|^2 + \text{Cost}(P, Q) \quad (6)$$

avec

- $X$  la matrice des ratings
- $P$  et  $Q$  les matrices des représentations latentes
- $\text{Cost}$  des contraintes sur  $P$  et  $Q$  :
  - Positivité (NMF)
  - Sparsité des représentations (de type  $L_1$ )
  - Dictionnaire over-complet (de taille très grande)
  - Simétrie
  - ....

# Factorisation de matrices et CF

$$\begin{matrix} \begin{matrix} 1 & 3 & 5 & 5 & 4 \\ 5 & 4 & 4 & 2 & 1 \\ 2 & 4 & 1 & 2 & 3 \\ 2 & 4 & 5 & 4 & 2 \\ 4 & 3 & 4 & 2 & 5 \end{matrix} & \sim & \begin{matrix} .1 & -.4 & .2 \\ -.5 & .6 & .5 \\ -.2 & .3 & .5 \\ 1.1 & 2.1 & .3 \\ -.7 & 2.1 & -2 \\ -1 & .7 & .3 \end{matrix} \\ \begin{matrix} 1.1 & -.2 & .3 & .5 & -.2 & -.5 & .8 & -.4 & .3 & 1.4 & 2.4 & -.9 \\ -.8 & .7 & .5 & 1.4 & .3 & -1 & 1.4 & 2.9 & -.7 & 1.2 & -.1 & 1.3 \\ 2.1 & -.4 & .6 & 1.7 & 2.4 & .9 & -.3 & .4 & .8 & .7 & -.6 & .1 \end{matrix} \end{matrix}$$

La SVD est indéfinie dans le cas de matrices incomplètes. On va donc utiliser d'autres algorithmes d'apprentissage :

- Stochastic Gradient Descent
- Optimisation alternée

Attention au **sur apprentissage** dans le cas où la matrice d'entrée est très sparse !!!  $\Rightarrow$  régularisation

## Exercice

Soit :

- $U$  le nombre de users
- $I$  le nombre d'items
- $R = (r_{u,i})$  la matrice des ratings, et  $C = (c_{u,i})$  le masque associé tel que :

$$c_{u,i} \begin{cases} = 1 & \text{si l'utilisateur } u \text{ a fourni un rating sur } i \\ = 0 & \text{sinon} \end{cases} \quad (7)$$

- ① Ecrire le problème d'optimisation correspondant à une factorisation matricielle **avec contrainte  $L_2$**  qui ne prennent en compte uniquement les ratings fournis par les utilisateurs.
- ② Ecrire l'algorithme de SGD correspondant

# Correction

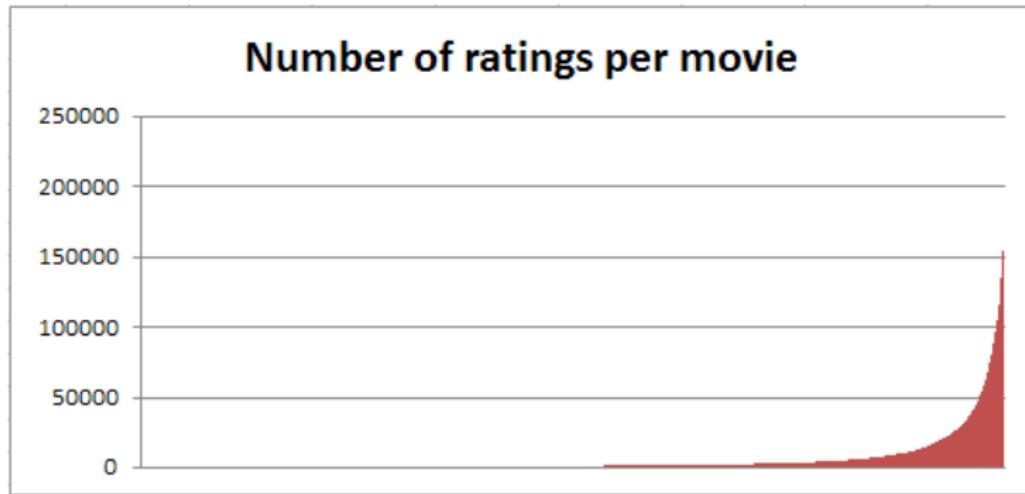
Au tableau

# Biais

## Problèmes

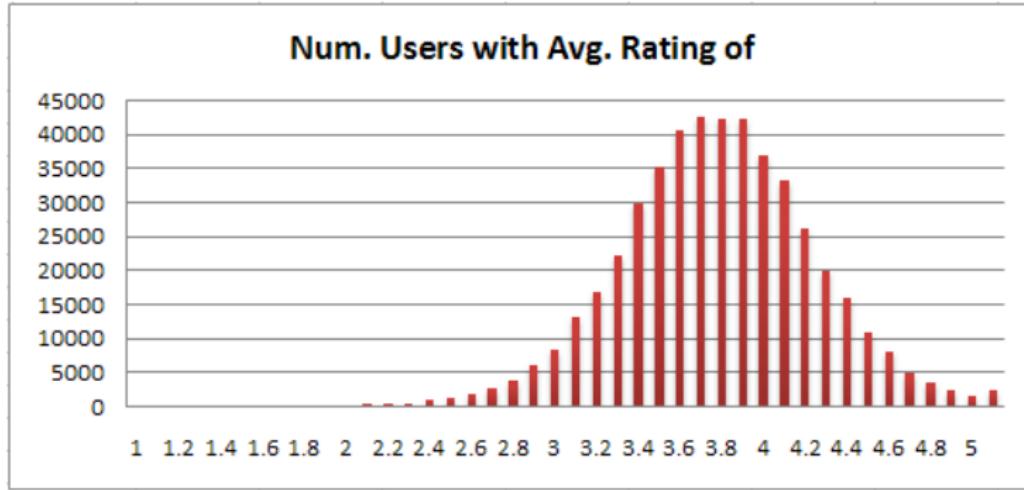
- Certains items sont toujours bien notés - Biais sur les items
- Certains utilisateurs sont optimistes (ou pessimistes) - Biais sur les utilisateurs
- Les notes changent en fonction de l'époque et du temps - Biais temporel

# Biais



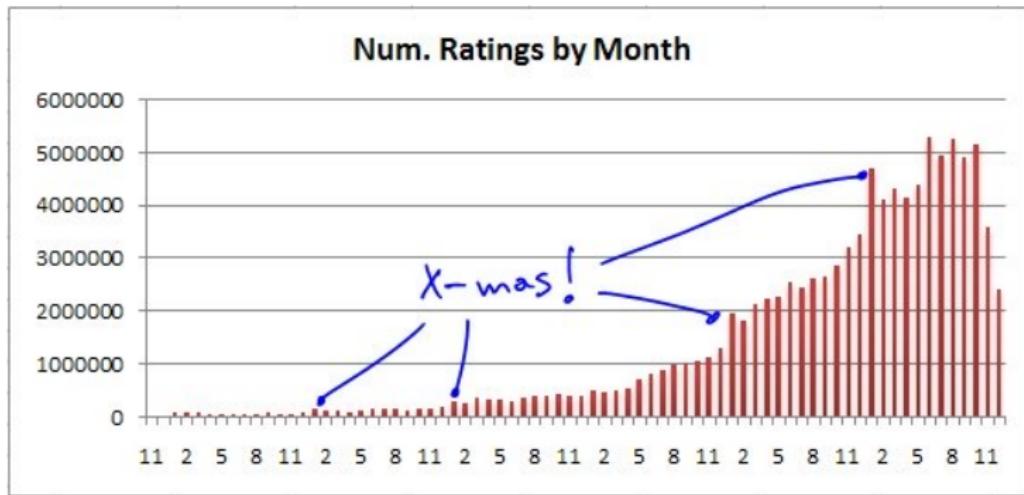
Source : <https://www.igvita.com/2006/10/29/dissecting-the-netflix-dataset/>

# Biais



Source : <https://www.igvita.com/2006/10/29/dissecting-the-netflix-dataset/>

# Biais



Source : <https://www.igvita.com/2006/10/29/dissecting-the-netflix-dataset/>

## Exercice

Comment intégrer le bias utilisateur et le biais sur les items ?

- Chaque item a une note moyenne (qui ne dépend pas de l'utilisateur)
- Chaque utilisateur est associé à une note moyenne (qui ne dépend pas de l'item)

# Intégration du biais utilisateur et items

$$\begin{aligned} \text{minimize}_{p,q} \quad & \sum_{(u,i) \in S} (r_{ui} - (\mu + b_u + b_i + \langle p_u, q_i \rangle))^2 + \\ & \lambda \left[ \|p\|_{\text{Frob}}^2 + \|q\|_{\text{Frob}}^2 + \|b_{\text{users}}\|^2 + \|b_{\text{items}}\|^2 \right] \end{aligned}$$

Descente de gradient stochastique :

$$p_u \leftarrow (1 - \lambda \eta_t) p_u - \eta_t q_i \rho_{ui}$$

$$q_i \leftarrow (1 - \lambda \eta_t) q_i - \eta_t p_u \rho_{ui}$$

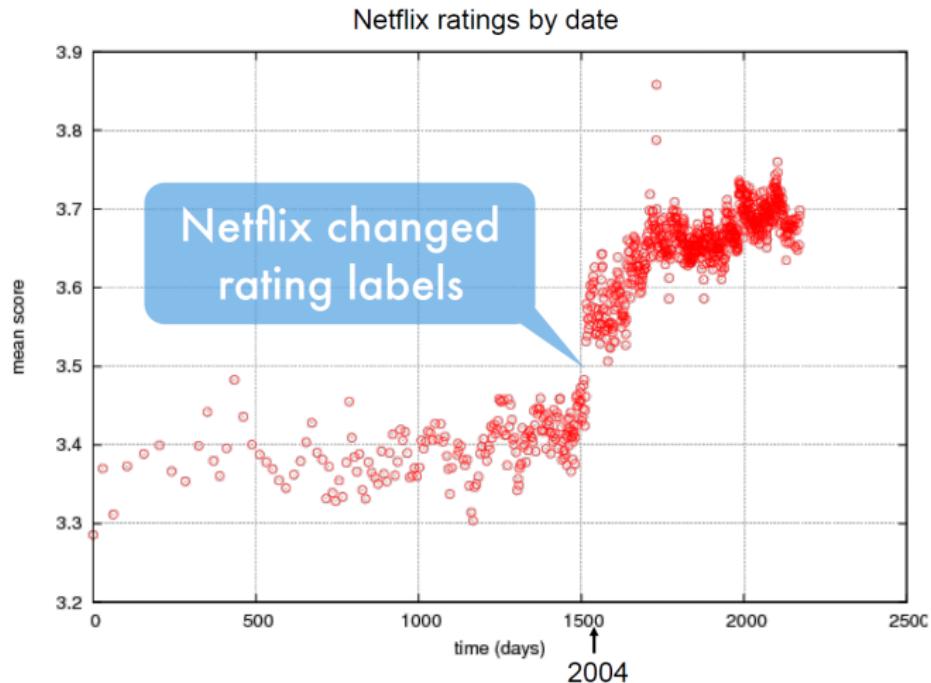
$$b_u \leftarrow (1 - \lambda \eta_t) b_u - \eta_t \rho_{ui}$$

$$b_i \leftarrow (1 - \lambda \eta_t) b_i - \eta_t \rho_{ui}$$

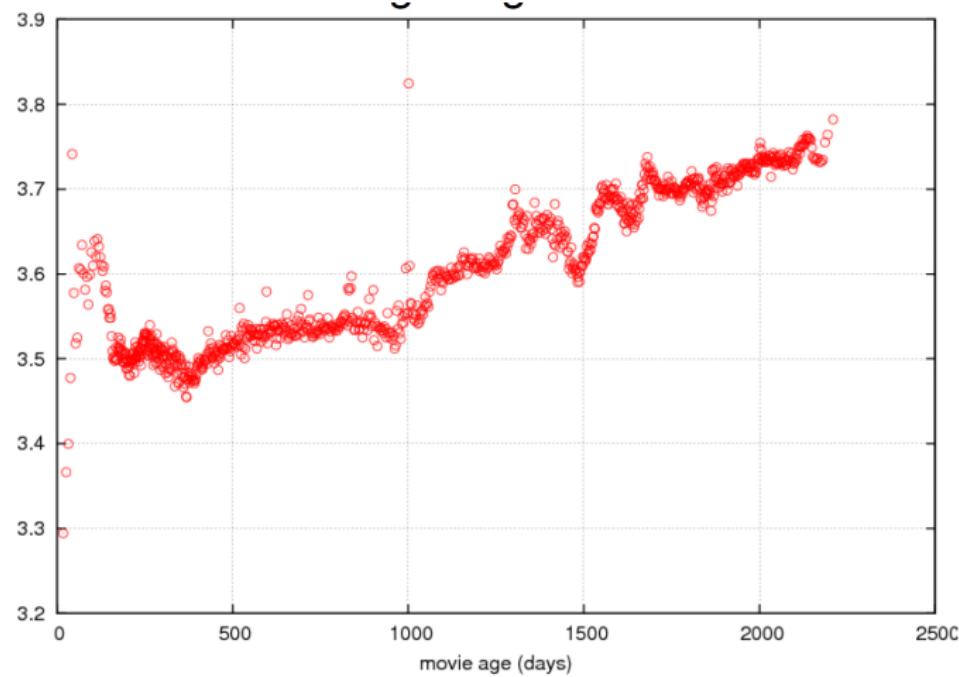
$$\mu \leftarrow (1 - \lambda \eta_t) \mu - \eta_t \rho_{ui}$$

where  $\rho_{ui} = (r_{ui} - (\mu + b_i + b_u + \langle p_u, q_i \rangle))$

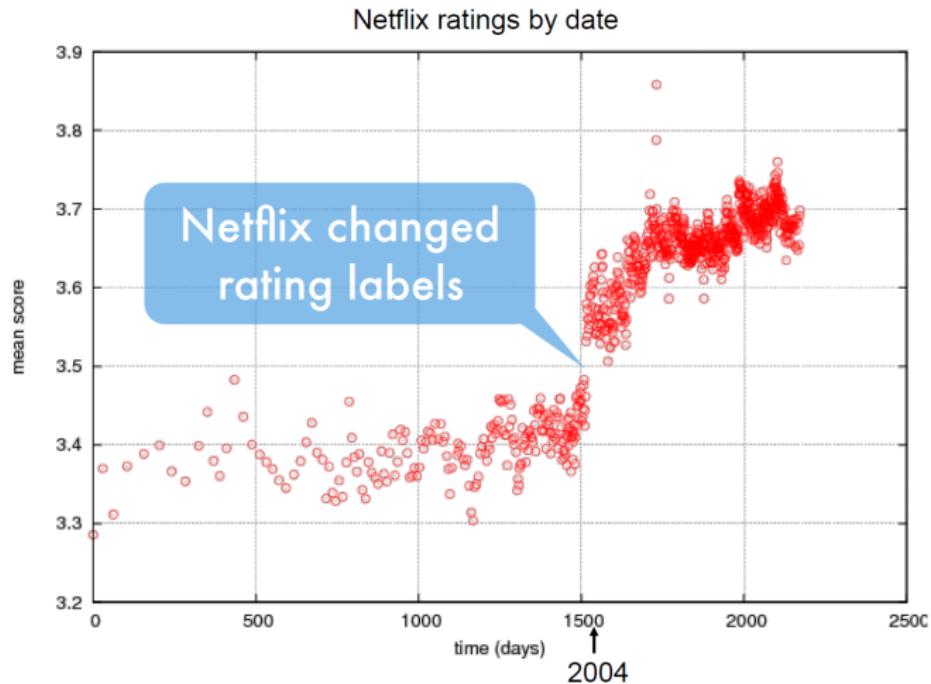
# Intégration du biais temporel



# Intégration du biais temporel



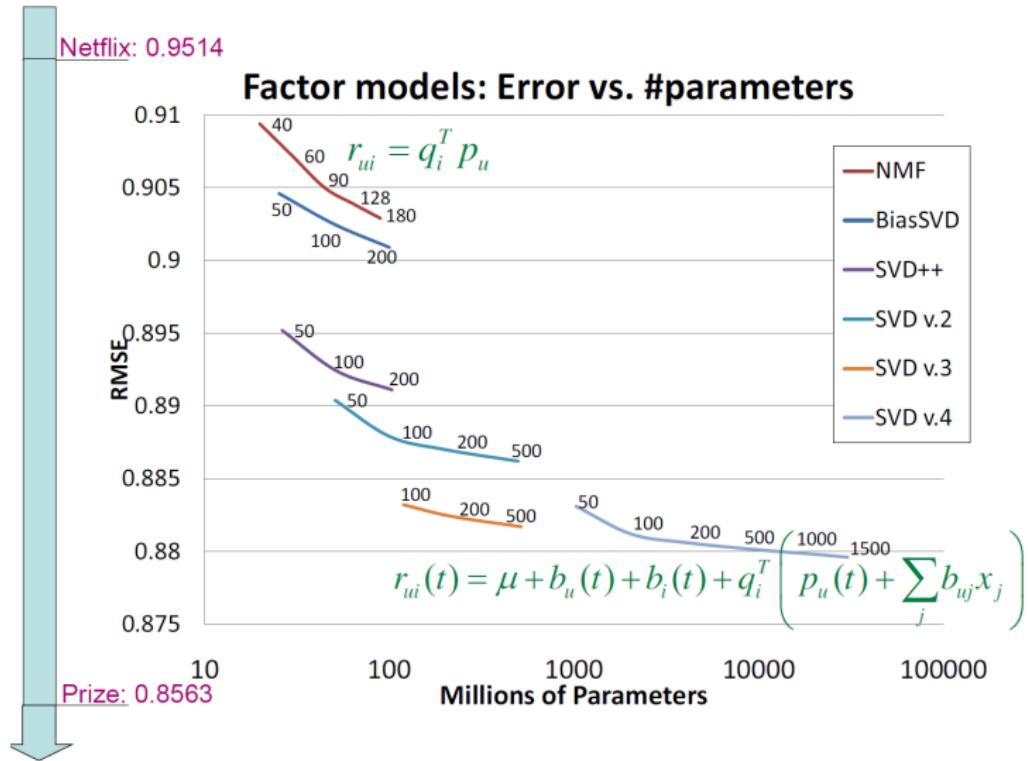
# Intégration du biais temporel



## Exercice :

Comment intégrer le biais temporel dans le modèle précédent ?

# Au final



# Conclusion