

Factorisation matricielle

Application à la recommandation personnalisée de préférences

Julien Delporte

INSA de Rouen

3 février 2014

1 Les Systèmes de Recommandation

2 La Factorisation Matricielle

3 Recommandation sociale

4 Sélection de modèle

1 Les Systèmes de Recommandation

- Qu'est ce que c'est ?
- Des exemples
- Problématiques

Qu'est ce que c'est ?

amazon.fr

Chez vous | Promotions | Chèques-cadeaux | Vendre | Aide

Parcourir les boutiques ▼

Rechercher

DVD & Blu-ray ▼

Go

Bonjour. Identifiez-vous
Votre compte ▼

DVD & Blu-ray

Recherche détaillée

Nos rubriques

Meilleures ventes

Nouveautés et À paraître

Promotions

Blu-ray

Séries TV

Jeunesse

Vendez !

Retourner ou remplacer une commande cadeau Fenêtre de retour
étendue jusqu'au 31 janvier 2014

> Rendez-vous dans notre Centre de retours



DVD & Blu-ray : accès rapides

A ne pas manquer

Satisfait ou remboursé jusqu'au 31 Janvier 2014

- 5 DVD = 30€

- 50€ de réduction dès 100€ d'achats de DVD, Blu-ray & Séries TV

- 2 Blu-ray achetés = le 3e offert

- 5 Blu-ray achetés = 5 Blu-ray offerts

- 1 Série TV achetée = la 2ème à 10€

- Soldes et petits prix DVD & Blu-ray

> Toutes les promos

Formats

DVD

Blu-ray

Blu-ray 3D active

Coffrets collectors

Sous-titres pour malentendants en français

> Tous les formats

Genres

Action, Policier et Thriller

Jeunesse et Famille

Séries TV

Films musicaux et concerts

Drame et Émotion

Comédie

Fantastique, Horreur et Science-fiction

Cinéma Asiatique

Documentaires et Divers

Grands classiques

DVD & Blu-ray

DVD | Blu-ray | Séries TV | Jeunesse | Disney | Nouveautés | Promotions Pour les offres ci-dessous, voir conditions sur la page relayant l'offre.

A ne pas manquer



No Limit - Saison 2
Vincent Elbaz, Anne Girouard, ...
DVD
EUR-29.06 EUR 17,99



Diana
Naomi Watts, Naveen Andrews, ...
DVD
★★★★☆ (2)
EUR 19,99



L'Intégrale "Le Bôcoup Meilleur"
Inconnus Les
DVD
★★★★☆ (70)
EUR 19,99



South Park - Saison 16
Trey Parker, Matt Stone, Jesse
Howell...
DVD
EUR 29,99

En lien avec des articles que vous avez regardés

Vous avez regardé



Stanley Kubrick: Collection :
Shining
Jack Nicholson, Shelley Duvall, ...
DVD
★★★★☆ (80)
EUR-19.02 EUR 9,02



Vol au-dessus d'un nid de coucou
Jack Nicholson, Louise Fletcher, ...
DVD
★★★★☆ (45)
EUR 9,99



Kubrick: réalisateur visionnaire...
Stanley Kubrick
Blu-ray
★★★★☆ (8)
EUR-70.02 EUR 33,99

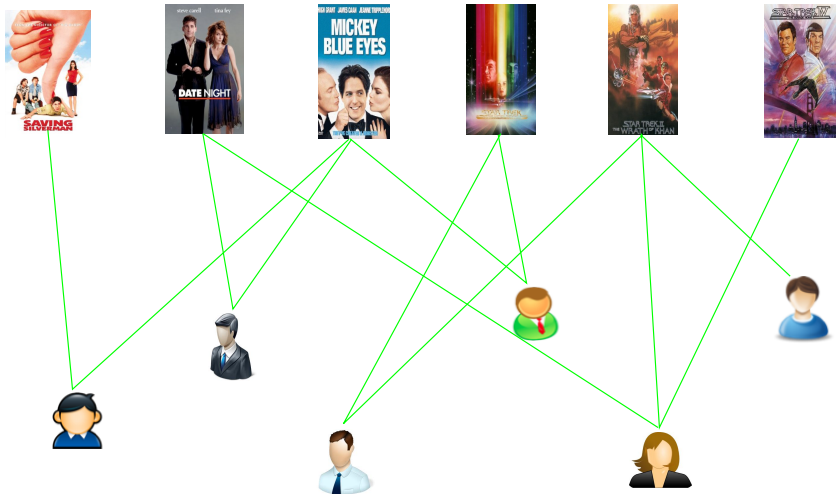


Stanley Kubrick: Collection :
Full...
Matthew Modine, Adam Baldwin, ...
DVD
★★★★☆ (52)
EUR 7,82

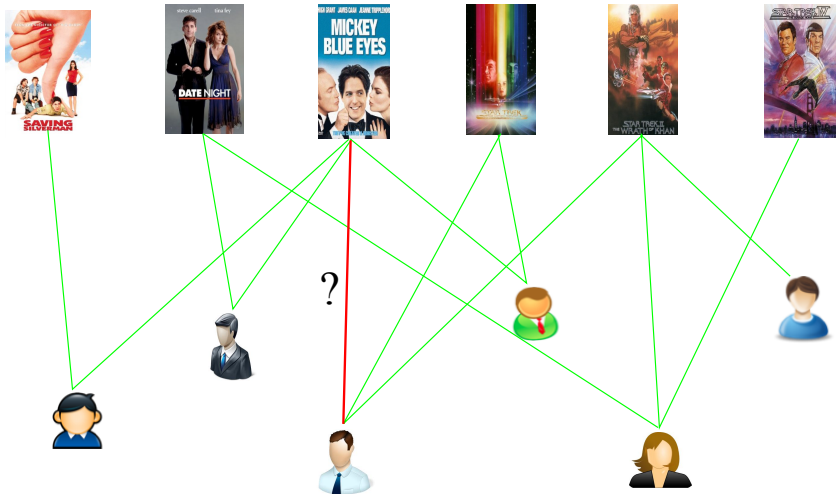
Quelques exemples



Prédiction de préférences

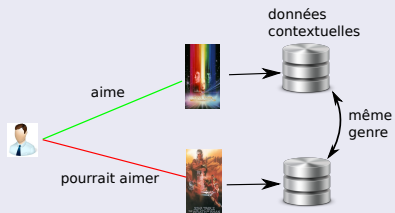


Prédiction de préférences

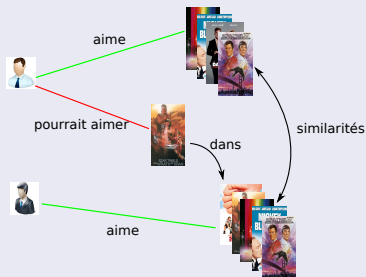


Méthodes de prédictions existantes

Approche orientée contenu



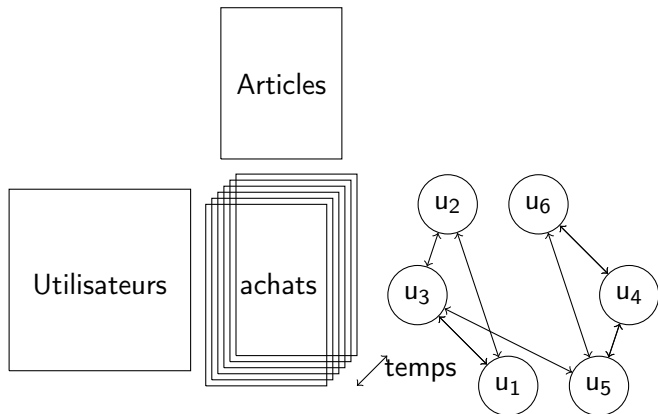
Filtrage collaboratif



Approche hybride

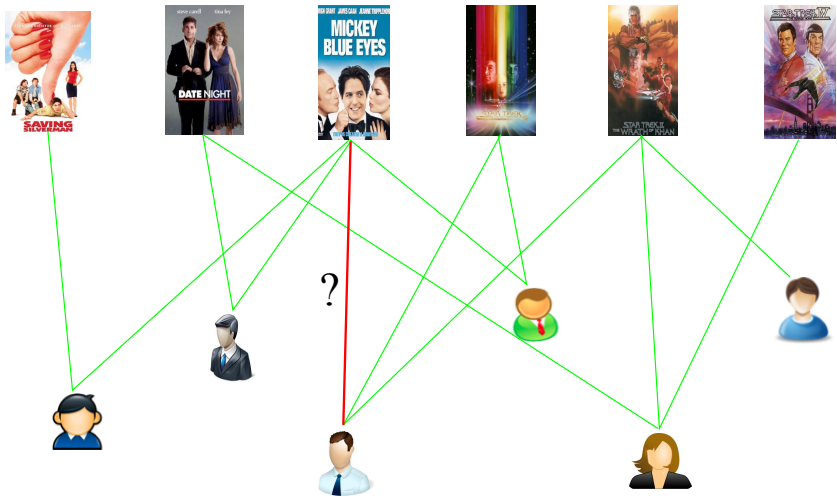
Recommande à partir des données de préférences, et d'autres informations dites contextuelles

Les difficultés

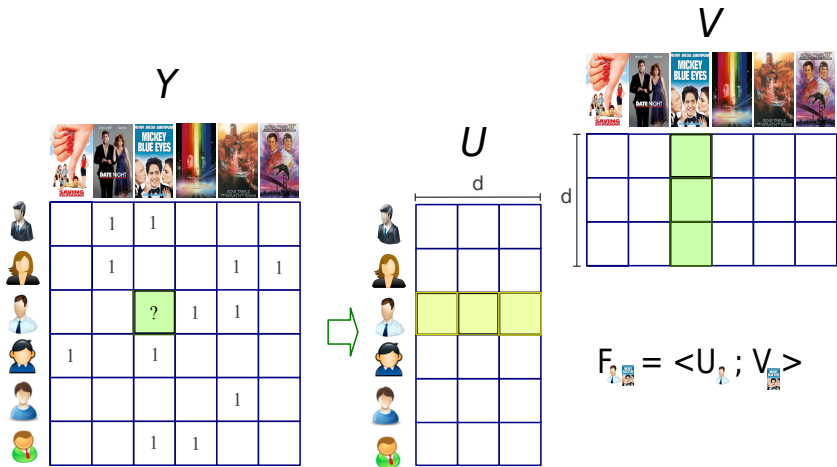


- 2 La Factorisation Matricielle
 - Problématique générale
 - Formalisation du problème
 - Méthodes de factorisation existantes
 - Bilan

Factorisation pour la recommandation



Factorisation pour la recommandation



Formalisation du problème

Modèle

$$Y = M + R$$

avec M régulière et R une matrice de bruit.

Formalisation du problème

Modèle

$$Y = M + R$$

avec M régulière et R une matrice de bruit.

Problème

$$\min_F \mathcal{L}(Y, F) + \lambda \Omega(F)$$

avec $F = UV$ une factorisation de faible rang

Le critère d'attache aux données \mathcal{L}

approche probabiliste

$$\mathcal{L}(Y, F) = - \sum_{(i,j)} \log P(Y_{ij}|F_{ij})$$

exemples

$$Y_{ij} \sim \mathcal{N}(M_{ij}, \sigma^2) \rightarrow \mathcal{L}(Y, F) = \sum_{(i,j)} (F_{ij} - Y_{ij})^2$$

$$Y_{ij} \sim \mathcal{B}(M_{ij}) \rightarrow \mathcal{L}(Y, F) = \sum_{(i,j)} \log(1 + \exp(-Y_{ij}F_{ij}))$$

Le terme de pénalisation Ω

norme visant à

- contrôler le rang ($\|\cdot\|_*$, $rg(\cdot)$)
- induire la parcimonie ($\|\cdot\|_0$, $\|\cdot\|_1$)
- de façon générale à pénaliser la complexité du modèle.

Méthodes existantes

Décomposition en valeurs singulières

Problème :

$$\min_F \|Y - F\|_{Fro}^2$$

$$s.t. \quad rang(F) = k$$

Solution (Eckart & Young, 1936) :

$$F = U_{(k)} \Sigma_{(k)} V_{(k)}^T$$

Algorithme : Processus de Lanczos (Larsen, 1998)

Méthodes existantes

Méthode additive

- $U_i \leftarrow U_i - \rho \nabla_{U_i} J(Y, F)$
- $V_j \leftarrow V_j - \rho \nabla_{V_j} J(Y, F)$

Avantage : adaptabilité, rapidité

Inconvénient : paramétrage difficile, convergence

Méthodes existantes

Méthode multiplicative

- $U \leftarrow U \otimes \frac{\nabla_U^- J(Y, F)}{\nabla_U^+ J(Y, F)}$
- $V \leftarrow V \otimes \frac{\nabla_V^- J(Y, F)}{\nabla_V^+ J(Y, F)}$

Avantage : contrainte de positivité (NMF), convergence

Inconvénient : adaptabilité, complexité

Méthodes existantes

Méthode multiplicative

- $U \leftarrow U \otimes \frac{\nabla_U^- J(Y, F)}{\nabla_U^+ J(Y, F)}$
- $V \leftarrow V \otimes \frac{\nabla_V^- J(Y, F)}{\nabla_V^+ J(Y, F)}$

Avantage : contrainte de positivité (NMF), convergence

Inconvénient : adaptabilité, complexité

exemple

- $J = \frac{1}{2} \|Y - UV\|_F^2 + \lambda(\|U\|^2 + \|V\|^2)$
- $\nabla_U J = -(Y - UV)V^T + 2\lambda I$
- $\nabla_U^+ J = UVV^T + 2\lambda I$
- $\nabla_U^- J = YV^T$

Méthodes existantes

Méthode analytique

- mettre à jour U_i en résolvant $0 = \nabla_{U_i} J(Y, F)$
- mettre à jour V_j en résolvant $0 = \nabla_{V_j} J(Y, F)$

Avantage : solution analytique, convergence

Inconvénient : adaptabilité, complexité

On fait le point

Les problèmes gérés par la factorisation

- Taille des données : oui
- Parcimonie : oui
- Variété des données : non
- nouveau problème : réglage du nombre de facteurs (d)

3 Recommandation sociale

- La notion de contextualisation
- Contexte social
- Méthodes existantes
- Notre modèle
- Expérimentations

La notion de contexte

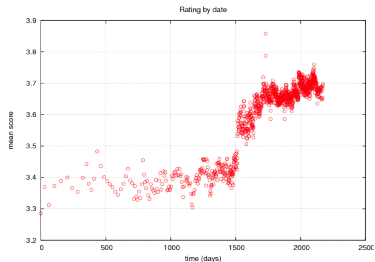
Définition

Tout type d'information qui peut être extraite de façon implicite ou explicite à partir des données disponibles.

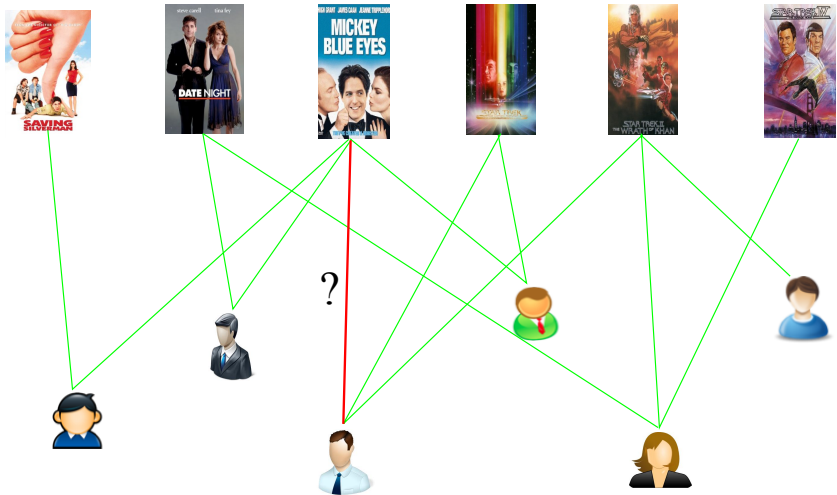
La notion de contexte

Des exemples de contexte

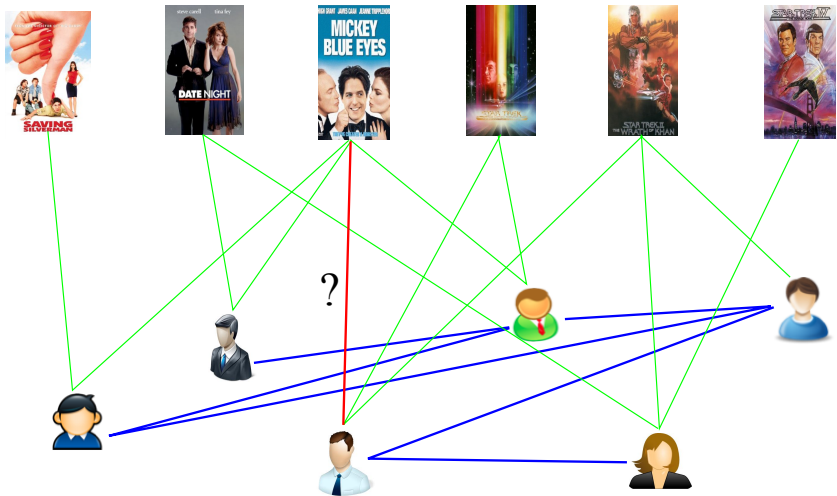
- temporel (saisonnalité, évolution comportementale...)
- démographique (âge, sexe, ville...)
- social
- ...



Contexte social





Contexte social



Notre problème

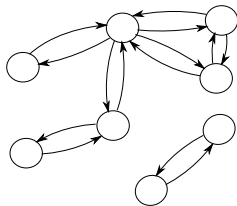
Taille des données

- 13M d'utilisateurs
- 100K endroits
- 200M connexions sociales

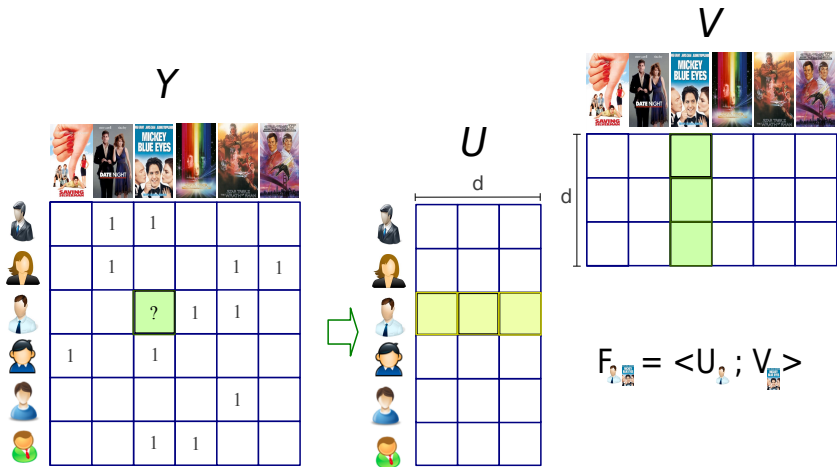
				...
user 1	1	0	1	...
user 2	1	1	0	...
user 3	0	0	1	...
user 4	0	1	0	...
⋮	⋮	⋮	⋮	

Parcimonie

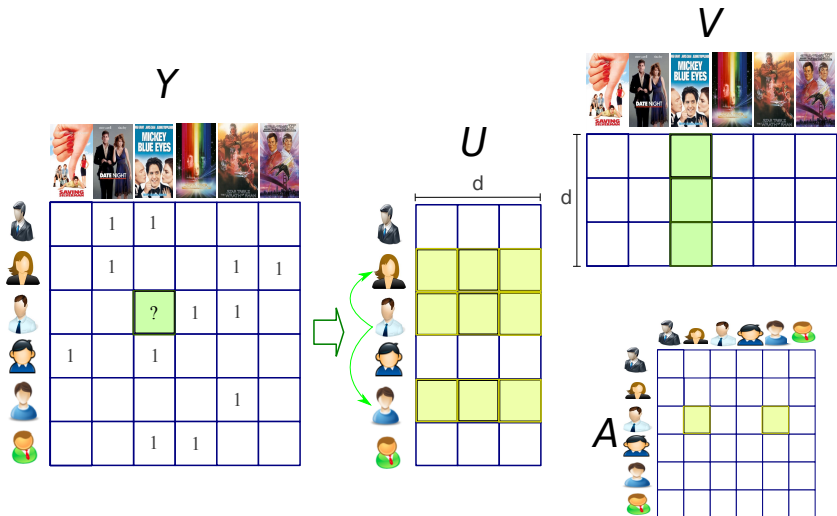
- 4 lieux par utilisateur
- 20 amis par utilisateur



Objectif



Objectif



Méthodes existantes

Modification de la fonction objectif

$$\min_{U,V} \mathcal{L}_1(Y, UV) + \mu \mathcal{L}_2(A, UU^T) + \lambda \Omega(U, V) \quad \text{Yang et al. (2011)}$$

$$\min_{U,V} \mathcal{L}(Y, UV) + \mu \sum_{i=1}^n \|U_i - \frac{1}{|\mathcal{F}_i|} \sum_{k \in \mathcal{F}_i} U_k\|_F^2 + \lambda \Omega(U, V) \quad \text{Ma et al. (2011)}$$

- \mathcal{F}_i l'ensemble des amis de i
- $A_{ij'} = 1$ si $i' \in \mathcal{F}_i$

Modification de la fonction de décision

$$\min_{U,V} \sum_{(i,j) \in \mathcal{Y}} c_{ij} \left(U_i V_j + \sum_{k \in \mathcal{F}_i} \frac{A_{ik} U_k V_j}{|\mathcal{F}_i|} - Y_{ij} \right)^2 + \Omega_{U,V,A} \quad \text{Ma et al. (2009)}$$

Apprentissage de l'influence

Fonction de décision

$$\hat{Y}_{ij} = U_i V_j + \sum_{k \in \mathcal{F}_i} \frac{A_{ik} U_k V_j}{|\mathcal{F}_i|} \quad \text{Ma et al. (2009)} \quad (1)$$

Fonction objectif

$$\min_{U, V, A} \sum_{(i,j) \in \mathcal{Y}} c_{ij} \left(U_i V_j + \sum_{k \in \mathcal{F}_i} \frac{A_{ik} U_k V_j}{|\mathcal{F}_i|} - Y_{ij} \right)^2 + \Omega_{U, V, A} \quad (2)$$

- c_{ij} un poids pour pénaliser les 1 plus que les 0.
- $\Omega_{U, V, A} = \lambda_1 \|U\|_{\text{Fro}}^2 + \lambda_2 \|V\|_{\text{Fro}}^2 + \lambda_3 \|A\|_{\text{Fro}}^2$

Contribution

Apprendre les influences des amis

Algorithmique

Solution analytique : moindres carrés alternés

Entrées : Y, \mathcal{F}

Initialiser U, V et A

Répéter

 Répéter

 Pour $i \in \mathcal{U}$ faire

 | mettre à jour U_i

 Fin Pour

 Pour $j \in \mathcal{V}$ faire

 | mettre à jour V_j

 Fin Pour

 jusqu'à ce que (U et V aient convergé)

 Pour $i \in \mathcal{U}$ faire

 Pour $i' \in \mathcal{F}_i$ faire

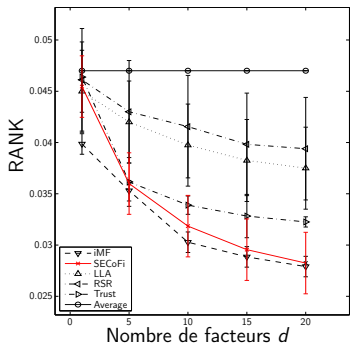
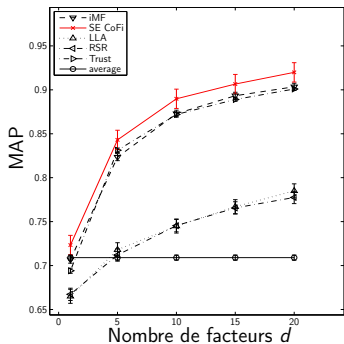
 | mettre à jour $A_{ii'}$

 Fin Pour

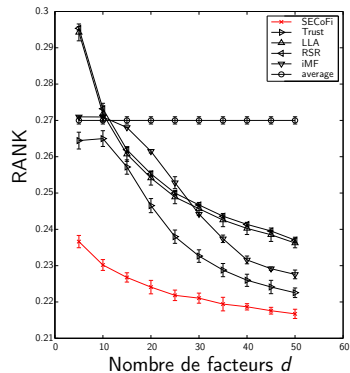
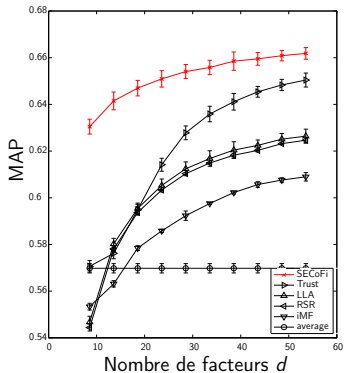
 Fin Pour

 jusqu'à ce que (convergence globale)

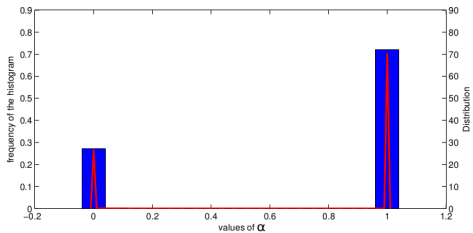
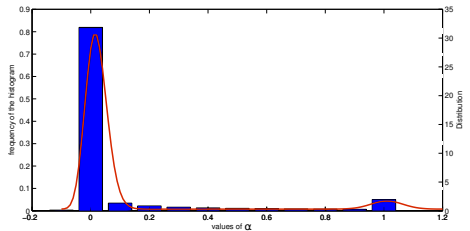
Résultats sur Tuenti



Résultats sur Epinions



A propos du graphe d'influence



Conclusion

Bilan de cette contribution

- Recommandation contextuelle sociale
- Apprentissage des influences

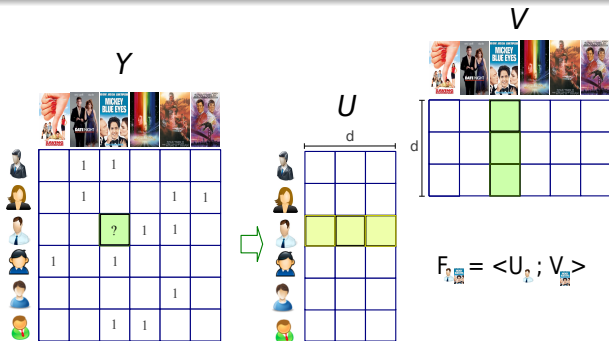
4 Sélection de modèle

- La sélection de modèle pour la factorisation
- Notre approche
- Élargissement du modèle
- Passage à l'échelle

Sélection de modèle pour la factorisation

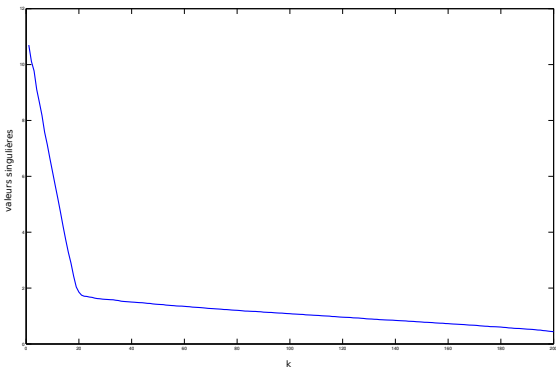
Objectif

Déterminer le nombre de facteurs adéquat pour la factorisation



Méthodes existantes

Méthode du coude



Méthodes existantes

Minimisation de risque

Modèle :

$$Y = M + R$$

Objectif :

$$\hat{M}^{(\lambda)} = f_{\lambda}(Y)$$

$$\min_{\lambda} \|Y - f_{\lambda}(Y)\|^2 + \text{pen}(\lambda, Y, f_{\lambda}(Y))$$

Notre point de départ

Stein (1977), Konno (1992), Candès (2012)

Soit $Y = U\Sigma V^\top$ et $f_\lambda(Y) = \sum_{k=1}^p f_k(\sigma_k) U_{\bullet k} V_{\bullet k}^\top$.

$$\min_{\lambda} \|Y - f_\lambda(Y)\|^2 + (2\text{div}_Y(f_\lambda(Y)) - np)\hat{s}^2$$

avec $f_k(x) = \max(0, x - \lambda)$.

Expression de la divergence

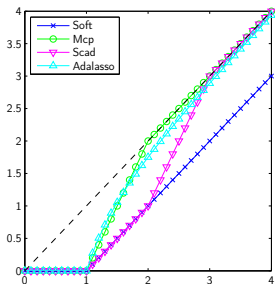
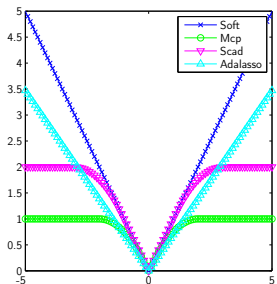
$$\text{div}_Y(f(Y)) = \sum_{i=1}^p \left(f'_i(\sigma_i) + (n-p) \frac{f_i(\sigma_i)}{\sigma_i} \right) + 2 \sum_{i=1}^p \sum_{\substack{j=1 \\ j \neq i}}^p \frac{\sigma_i f_i(\sigma_i)}{\sigma_i^2 - \sigma_j^2}$$

Notre point de départ

Limitations

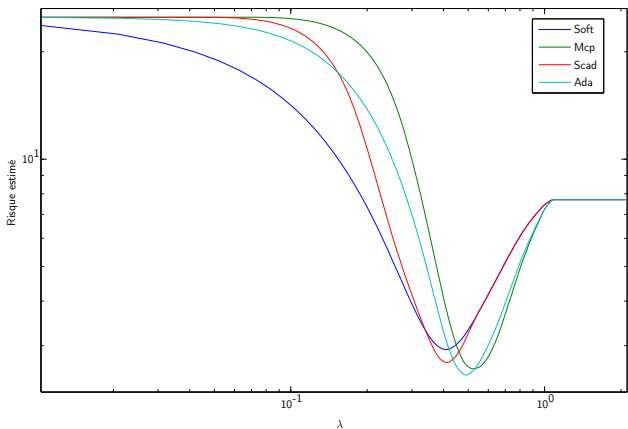
- a priori sur les données limité : utiliser d'autres fonctions f .
- div non calculable en grandes dimensions : approximer div .
- variance $\hat{\sigma}^2$ supposée connue : proposer une estimation.

Modification de la fonction de seuillage



fonctions de pénalité (gauche) et fonctions de seuillage (droite)

Modification de la fonction de seuillage



Estimation du risque en fonction du paramètre λ

Estimation de la divergence

Problème

$$\operatorname{div}_Y(f(Y)) = \sum_{i=1}^p \left(f'_i(\sigma_i) + (n-p) \frac{f_i(\sigma_i)}{\sigma_i} \right) + 2 \sum_{i=1}^p \sum_{\substack{j=1 \\ j \neq i}}^p \frac{\sigma_i f_i(\sigma_i)}{\sigma_i^2 - \sigma_j^2}$$

Solution proposée

- Calculer les k premières valeurs singulières.
- Estimer le reste du spectre.

Estimation de la divergence

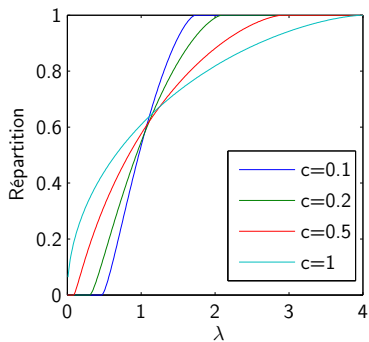
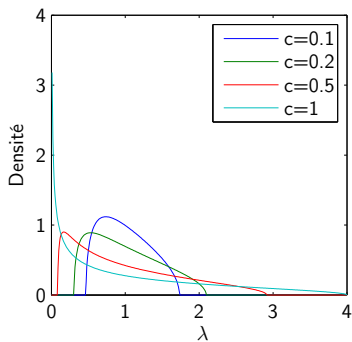
Théorème

Soit $X \in \mathcal{L}(p, n)$ une matrice aléatoire centrée. Les valeurs propres λ_i de la matrice $\frac{1}{n}X^\top X$ suivent (si on les considère comme variables aléatoires) une loi de Marchenko-Pastur de paramètre $c > 0$ dont la fonction de densité de probabilité vaut :

$$P_c(x) = \begin{cases} \frac{1}{2\pi xc} \sqrt{(x-a)(b-x)} & \text{si } a \leq x \leq b, \\ 0 & \text{sinon.} \end{cases}$$

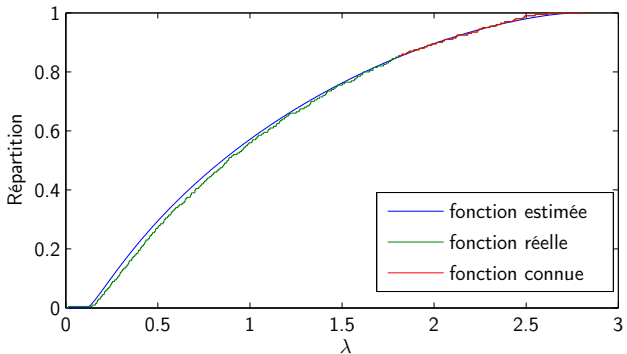
où $a = (1 - \sqrt{c})^2$ et $b = (1 + \sqrt{c})^2$.

Estimation de la divergence



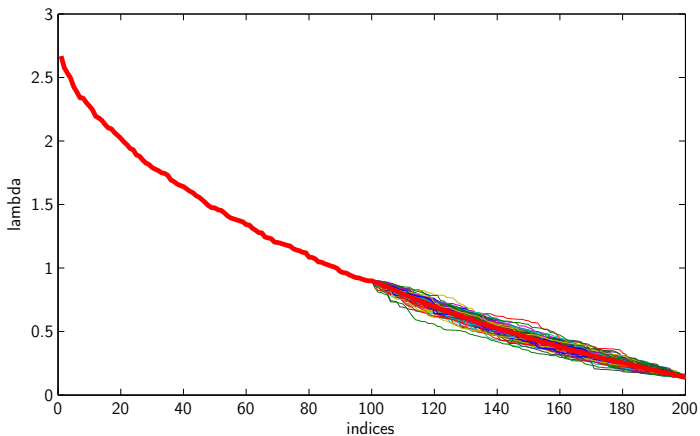
Loi de Marchenko-Pastur

Estimation de la divergence



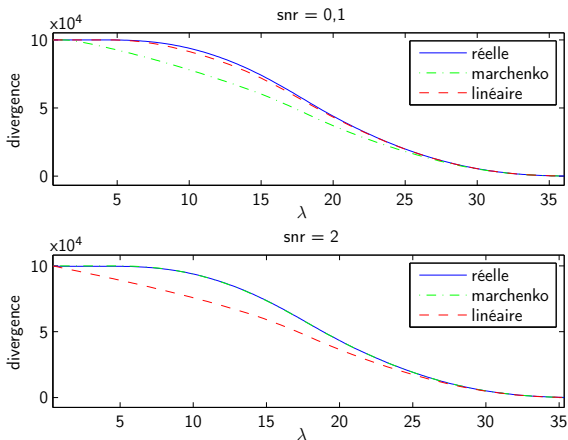
Approximation de la loi de Marchenko-Pastur

Estimation de la divergence



Tirages selon la loi de Marchenko-Pastur

Estimation de la divergence



Comparaison d'estimations de divergence

Conclusion

Recommandation sociale

- Apprentissage d'influences
- Perspectives : Recommandation d'amis

Sélection de modèle

- Passage à l'échelle et Élargissement
- Perspectives : Test réel, généraliser l'hypothèse du bruit

Conclusion

Communications

- Bruno Pradel, Savaneary Sean, Julien Delporte, Sébastien Guérif, Céline Rouveirol, Nicolas Usunier, Françoise Fogelman-Soulié et Frédéric Dufau-Joel. *A case study in a recommender system based on purchase data*. SIGKDD'11, pages 377–385. ACM, 2011.
- Julien Delporte, Alexandros Karatzoglou, Tomasz Matuszczyk et Stéphane Canu. *Socially Enabled Preference Learning from Implicit Feedback Data*. ECML PKDD'13, pages 145–160. Springer, 2013. (Best Student Paper)
- Julien Delporte, Alexandros Karatzoglou et Stéphane Canu. *Apprentissage et factorisation pour la recommandation*. RNTI, vol. RNTI-A-6, 2014.