

LOG6308 — Systèmes de recommandation

Systèmes de recommandations 3, Factorisation et réduction de dimensions

Michel C. Desmarais

Génie informatique et génie logiciel
École Polytechnique de Montréal

Automne, 2017
(version 19 septembre 2017)

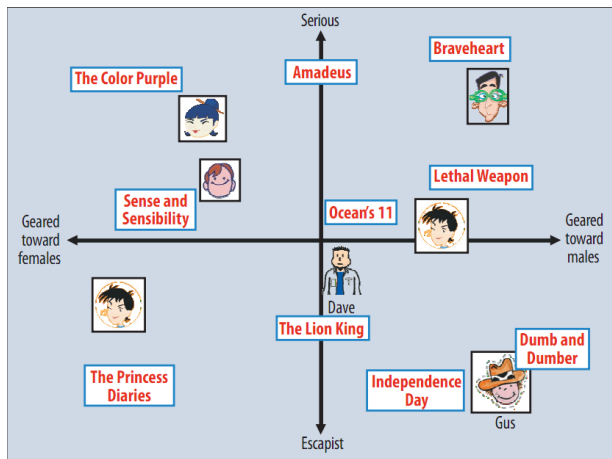
Systèmes de recommandations 3, Factorisation et réduction de dimensions

- 1 Introduction
- 2 Décomposition en valeurs singulières
- 3 Autres factorisation et valeurs manquantes
- 4 Valeurs manquantes

Factorisation et réduction de dimensions

- La factorisation permet de représenter une matrice par le produit de deux autres matrices
 - $\mathbf{M}_{(m,n)} = \mathbf{Q}_{(m,l)} \mathbf{P}_{(n,l)}^T$
 - La dimension commune, l , représente les facteurs latents
- Cette représentation constitue une forme de compression des données
 - Si l'égalité tient, au lieu de $m \times n$ valeurs, on a $m \times l + n \times l$
- De très nombreuses applications
 - Compression d'images, analyse sémantique texte, évaluation cognitive, etc.

Films et utilisateurs autour de 2 dimensions



2 films, 2 utilisateurs, de 2 dimensions

Deux films, deux individus :

	Princess	Orange
Rosa R.	2	-2
Stanly K.	-2	2



- L'idée est de partir de cette matrice pour dériver les matrices film (item) et individu (utilisateur). *N.B. Ignorons pour le moment la singularité de la matrice.



2 films, 2 utilisateurs, de 2 dimensions

- Supposons deux dimensions de films :

- Humeur** : léger \longrightarrow sérieux
- Genre** : féminin \longrightarrow masculin

- Matrice films :

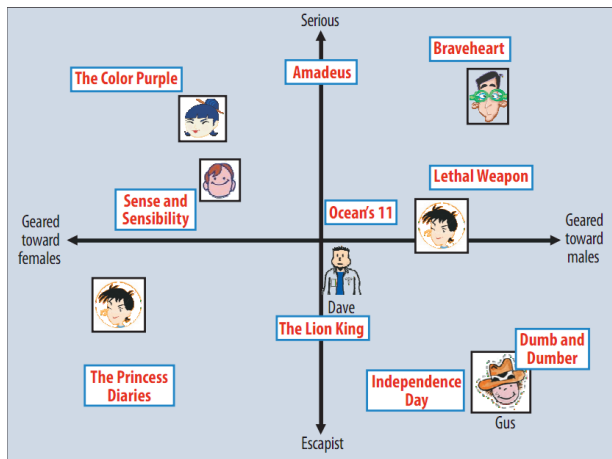
Film	Humeur	Genre
The Princess Diaries	léger (-1)	fém. (-1)
Orange mécanique	sérieux (+1)	masc. (+1)

- Matrice utilisateur :

Individu	Humeur	Genre
Rosa Rose	-1	-1
Stanly Kubrick	+1	+1



Films et utilisateurs autour de 2 dimensions



Principes

$$\mathbf{M}_{m \times n} = \mathbf{Q}_{m \times l} \mathbf{P}_{n \times l}^T$$

- Plusieurs techniques de décomposition :
 - On factorise à partir des vecteurs propres de $\mathbf{M}\mathbf{M}^T$ ce qui entraîne l'orthogonalité (analyse en **composante principales**, PCA)
 - Dans le cas de la technique **SVD**, on impose aussi la normalité (vecteurs unitaires)
 - La factorisation selon les **valeurs non négatives** (NMF) apporte des solutions parfois plus naturelles à interpréter
- Prédiction à partir de la factorisation
 - La factorisation produit une **matrice dense**, de laquelle des calculs de **similarité** peuvent être effectués
 - La factorisation représente un modèle et le produit de \mathbf{Q} et \mathbf{P} constitue une estimation (prédiction), $\hat{\mathbf{M}}$

Factorisation de matrice pour filtres collaboratifs

Supposons la matrice utilisateurs-items \mathbf{M} , il s'agit alors de la factoriser en deux matrices \mathbf{Q} et \mathbf{P} :

$$\mathbf{M} = \mathbf{Q}\mathbf{P}^T$$

- Cette approche permet d'identifier des **facteurs latents** qui “expliquent” le vote
 - Les facteurs se combinent linéairement pour augmenter/diminuer le vote (somme)
- \mathbf{Q} représente la préférence des utilisateurs pour les facteurs
- \mathbf{P} représente l'appartenance des items aux facteurs
- **A donné les meilleurs résultats à Netflix en 2009**

Techniques de factorisation

- **SVD**
 - Méthode déjà bien connue en recherche d'information
- **Décomposition en valeurs positives**
 - N'exige pas l'orthogonalisation de \mathbf{Q}_i et \mathbf{P}_u ce qui peut donner des interprétations plus naturelles
 - Solutions multiples
- Il existe différents algorithmes pour converger vers une solution

Systèmes de recommandations 3, Factorisation et réduction de dimensions

- 1 Introduction
- 2 Décomposition en valeurs singulières
- 3 Autres factorisation et valeurs manquantes
- 4 Valeurs manquantes

Décomposition en valeurs singulières I

$$\mathbf{M} = \mathbf{U}\mathbf{\Sigma}\mathbf{V}^T$$

- Décomposition en trois matrices :

U : matrice orthonormale qui représente les vecteurs propres de $\mathbf{M}\mathbf{M}^T$

V : matrice orthonormale qui représente les vecteurs propres de $\mathbf{M}^T\mathbf{M}$

Σ : matrice diagonale qui représente les *valeurs singulières* de **M** (les racines des valeurs propres des vecteurs propres de $\mathbf{M}\mathbf{M}^T$ et $\mathbf{M}^T\mathbf{M}$.)

Décomposition en valeurs singulières II

- Les valeurs singulières de Σ sont ordonnées et indiquent l'amplitude relative de chaque vecteur propre.
- On utilise les n premières valeurs singulières de Σ pour effectuer une réduction de dimension qui peuvent représenter des facteurs latents. C'est ce que l'on nomme la **LSA (Latent Semantic Analysis)**¹.

1. La communauté de RI y réfère par LIS (*Latent Semantic Indexing*). 

Exemple dans le domaine de la recherche d'information

- Le domaine de la **recherche d'information** (RI) partage plusieurs techniques avec les systèmes de recommandations (en fait, ces techniques sont directement utilisables pour les approches contenu)
- La matrice en RI est plutôt de la forme :

$$\begin{array}{c} \text{documents} \\ \text{termes} \end{array} \left(\begin{array}{ccc} a_{11} & \dots & a_{1n} \\ \vdots & \ddots & \vdots \\ a_{m1} & \dots & a_{mn} \end{array} \right)$$

Exemple de contenu

(tiré de Landauer et al. 1998)

- c1 : *Human* machine *interface* for ABC *computer* applications
- c2 : A *survey* of *user* opinion of *computer system response time*
- c3 : The *EPS user interface* management *system*
- c4 : *System* and *human system* engineering testing of *EPS*
- c5 : Relation of *user* perceived *response time* to error measurement
- m1 : The generation of random, binary, ordered *trees*
- m2 : The intersection *graph* of paths in *trees*
- m3 : *Graph minors* IV : Widths of *trees* and well-quasi-ordering
- m4 : *Graph minors* : A *survey*

Matrice termes documents

	c1	c2	c3	c4	c5	m1	m2	m3	m4
human	1	0	0	1	0	0	0	0	0
interface	1	0	1	0	0	0	0	0	0
computer	1	1	0	0	0	0	0	0	0
user	0	1	1	0	1	0	0	0	0
system	0	1	1	2	0	0	0	0	0
response	0	1	0	0	1	0	0	0	0
time	0	1	0	0	1	0	0	0	0
EPS	0	0	1	1	0	0	0	0	0
survey	0	1	0	0	0	0	0	0	1
trees	0	0	0	0	0	1	1	1	0
graph	0	0	0	0	0	0	1	1	1
minors	0	0	0	0	0	0	0	1	1

La réduction de dimensions

Les **corrélations initiales** des paires de termes :

- *human-user* : -0,38
- *human-minors* : -0,29

Réduction des dimensions :

- On ne conserve que les deux premières valeurs singulières (sachant qu'il n'y a que deux sources de documents).
- La matrice originale est reproduite avec les dimensions réduites à 2 (toutes les valeurs singulières de la matrice diagonale Σ sont mise à 0).

Les **corrélations** avec la nouvelle matrice sont maintenant :

- *human-user* : -0,94
- *human-minors* : -0,83

Ces chiffres sont bien davantage en ligne avec les attentes.

Réduction de dimensions

- La réduction de dimensions :
 - 1 permet d'assigner des valeurs sur une échelle continue là où il n'y avait que des valeurs $\{0,1\}$; elle offre donc une solution au problème des **matrices creuses** ;
 - 2 assigne des valeurs parfois très justes des appartenances des variables aux facteurs latents comme l'exemple de Landauer et al. le démontre.

Systèmes de recommandations 3, Factorisation et réduction de dimensions

- 1 Introduction
- 2 Décomposition en valeurs singulières
- 3 Autres factorisation et valeurs manquantes
- 4 Valeurs manquantes

Factorisation en valeurs non-négatives

$$\mathbf{R}_{u \times i} = \mathbf{W}_{u \times \ell} \mathbf{H}_{\ell \times i}$$

- Factorisation en deux matrices où ℓ représente les facteurs latents
- Contrainte de valeurs non-négatives
- Dans certains contexte, les valeurs négatives ne sont pas naturelles
- **Solutions multiples**
- Recherche par gradient
- On doit spécifier le nombre de **dimensions latentes** k

Factorisation excluant les valeurs manquantes et incluant les biais

(Koran, Bell et Volinsky, 2009)

$$\mathbf{R}_{u \times i} = \mathbf{P}_{u \times \ell} \mathbf{Q}_{\ell \times i}$$

- Pour l'ensemble des votes connus, κ , on cherche les valeurs de \mathbf{p} et \mathbf{q} qui minimisent l'équation :

$$\min_{\mathbf{p}, \mathbf{q}} \sum_{(u,i) \in \kappa} (r_{ui} - \mathbf{q}_i^T \mathbf{p}_u)^2 + \lambda (\|\mathbf{q}_i\|^2 + \|\mathbf{p}_u\|^2)$$

où $(\|\mathbf{q}_i\|^2 + \|\mathbf{p}_u\|^2)$ est un facteur de *régularisation* qui limite les valeurs de \mathbf{p}_u et \mathbf{q}_i et que l'on contrôle par λ .

Biais

Une proportion importante de la variance des votes s'explique par :

- le **biais utilisateur** : la propension d'un individu à des votes faibles ou forts
- le **biais item** : la popularité ou l'impopularité d'un item

Il faut donc inclure ces termes dans l'équation *objective* :

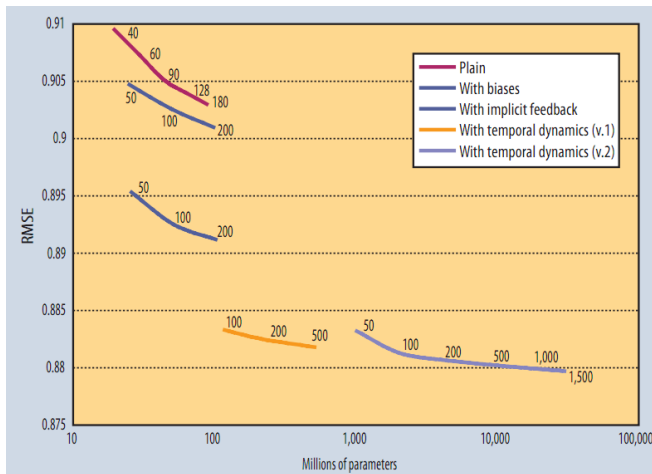
$$\hat{r}_{ui} = \mu + b_i + b_u + \mathbf{q}_i^T \mathbf{p}_u$$

où b_i et b_u sont les écarts à la moyenne pour l'item i et l'utilisateur u respectivement.

Autres facteurs

- Autres termes qui s'ajoutent à l'équation objective :
 - Votes implicites (traces utilisateur)
 - Facteurs démographiques
- Certains termes sont fonction du temps
 - $b_i(t), b_u(t)$
- Un vote peut avoir plus ou moins d'importance
 - c_{ui} est alors un facteur multiplicatif de \hat{r}_{ui}

Impact des facteurs (Koren et al., 2009)



0,951 était la performance du système Netflix initial, et **0,856** est la performance requise pour gagner le prix.

Systèmes de recommandations 3, Factorisation et réduction de dimensions

- 1 Introduction
- 2 Décomposition en valeurs singulières
- 3 Autres factorisation et valeurs manquantes
- 4 Valeurs manquantes

Factorisation avec les valeurs manquantes

- Typiquement les matrices de votes sont creuses.
- Deux solutions :
 - 1 Optimiser l'erreur quadratique avec la descente de gradient (sujet de projet)
 - 2 Estimer les valeurs manquantes afin de créer une matrice dense