

Factorisation matricielle non-négative sémantique

Mickael Febrissy¹ Aghiles Salah² Melissa Ailem³ Mohamed Nadif¹

¹LIPADE

UFR Mathématiques et Informatique
Université de Paris

²Singapore Management University

³University of Southern California

SFC'2019, Nancy



Table des matières

Factorisation matricielle non négative (NMF)

Factorisation matricielle non-négative sémantique (SNMF)

Expériences

Conclusion

La NMF, pour quels types données ?

Applicable sur des matrices de données non négatives. $\mathbf{X} \in \mathbb{R}_+^{n \times d}$

Matrices documents-termes (fouille textuelle)

- document 1 : La conférence se déroule à Nancy
- document 2 : Nancy est une ville Française
- document 3 : Paris est une ville Française

\mathbf{X}	conférence	Nancy	Française	ville
doc 1	1	1	0	0
doc 2	0	1	1	1
doc 3	1	0	1	1

Matrices utilisateurs-items (systèmes de recommandation)

	MindHunter	BlackMiroir	PeakyBlinders	Sherlock	Breaking Bad	The Wire
util 1	5	4	0	0	0	4
util 2	0	3	4	0	5	0
util 3	0	0	5	4	0	0
util 4	0	1	0	0	4	0
util 5	3	4	0	0	5	0

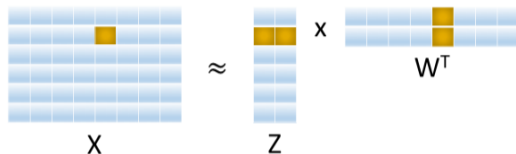
Des utilisateurs ayant noté des séries.

Autres

Données issues du domaine génétique, météorologique, financier, énergétique, etc.

Nonnegative Matrix Factorization (NMF)

Le modèle



$$\mathbf{X} \approx \mathbf{Z}\mathbf{W}^T \text{ avec } \mathbf{X} \in \mathbb{R}_+^{n \times d}, \mathbf{Z} \in \mathbb{R}_+^{n \times g} \text{ et } \mathbf{W} \in \mathbb{R}_+^{d \times g}$$

Caractéristiques

- Initialement une méthode de réduction de dimension
- \mathbf{Z} matrice des coefficients, \mathbf{W} matrice des centres
- g est choisi de façon à ce que $g(n + d) < ng$
- Sensibilité à l'initialisation.
- Références LEE et SEUNG 1999 ; LEE et SEUNG 2001

Factorisation matricielle non négative (NMF)

Fonctions objectives et algorithmes

- Norme de Frobenius : $\|\mathbf{X} - \mathbf{Z}\mathbf{W}^T\|_F^2$
- Divergences de Kullback-Leibler et Itakura-Saito.
- Algorithmes : règles multiplicatives, méthode des moindres carées (Alternative Least square), descente de gradient.
- Variantes probabilistes (Probabilistic NMF).

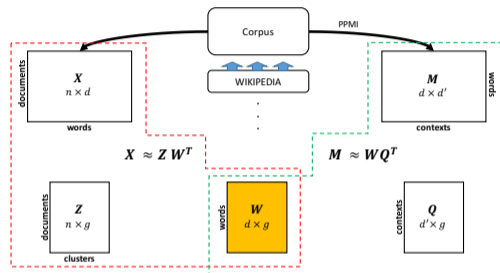
Cas particulier

Factorisation matricielle non négative symétrique (Symmetric NMF)

$$\mathbf{X} \approx \mathbf{Z}\mathbf{Z}^T \text{ avec } \mathbf{X} \in \mathbb{R}_+^{n \times n} \text{ et } \mathbf{Z} \in \mathbb{R}_+^{n \times g}$$

Semantic Nonnegative Matrix Factorization (SNMF)

Modèle



- D_1 et $D_2 \iff$ mesures d'erreur
- $\lambda \in \mathbb{R}_+$ un paramètre de régularisation.

$$\mathcal{F}(Z, W, Q) = \underbrace{D_1(X, ZW^T)}_{\text{NMF}} + \lambda \underbrace{D_2(M, WQ^T)}_{\text{word embedding}}$$

- $X \in \mathbb{R}_+^{n \times d}$, $Z \in \mathbb{R}_+^{n \times g}$, $W \in \mathbb{R}_+^{d \times g}$
- $M \in \mathbb{R}^{d \times d'}$, $Q \in \mathbb{R}^{d' \times g}$

Avec $D_1 = D_2 \iff$ la norme de Frobenius :

$$\mathcal{F}(Z, W, Q) = \frac{1}{2} \|X - ZW^T\|_F^2 + \frac{\lambda}{2} \|M - WQ^T\|_F^2 \quad (1)$$

Semantic Nonnegative Matrix Factorization (SNMF)

Matrice de contextes \mathbf{C}

$\mathbf{C} = (c_{jj'}) \in \mathbb{R}^{d \times d'}$ est une matrice de co-occurrences décrivant le nombre d'apparition pour chaque pair de termes $(t_j, t_{j'})$ d'un corpus de documents.

Pointwise Mutual Information (PMI)

Permet de quantifier l'association pair à pair entre les valeurs de deux variables discrètes aléatoires.

$$\text{PMI}(t_j, t_{j'}) = \log \frac{p(t_j, t_{j'})}{p(t_j)p(t_{j'})}$$

Suivant la nature de \mathbf{C} , la PMI peut se calculer ainsi par cette équation :

$$\text{PMI}(t_j, t_{j'}) = \log \frac{c_{jj'} \times c_{..}}{c_{j.} \times c_{.j'}}$$

où $c_{..} = \sum_{jj'} c_{jj'}$, $c_{j.} = \sum_{j'} c_{jj'}$ et $c_{.j'} = \sum_j c_{jj'}$. La PPMI (Positive PMI) est la version prenant en considération que les valeurs positives de la PMI (les valeurs négatives étant mises à 0).

$$\mathbf{M} = (m_{jj'} = \text{PPMI}(t_j, t_{j'})) \in \mathbb{R}^{d \times d'}$$

Algorithmme

Algorithm 1 NMF-Sémantique (SNMF).

Entrées : $\mathbf{X} \in \mathbb{R}_+^{n \times d}$, $\mathbf{M} \in \mathbb{R}_+^{d \times d'}$, λ et g .

Sorties : $\mathbf{Z} \in \mathbb{R}_+^{n \times g}$, $\mathbf{W} \in \mathbb{R}_+^{d \times g}$ et $\mathbf{Q} \in \mathbb{R}_+^{d' \times g}$.

Étapes :

1. Initialisation : $\mathbf{Z} \leftarrow \mathbf{Z}^{(0)}$; $\mathbf{W} \leftarrow \mathbf{W}^{(0)}$ et $\mathbf{Q} \leftarrow \mathbf{Q}^{(0)}$;

répéter

$$2. \mathbf{Z} \leftarrow \mathbf{Z} \odot \frac{\mathbf{X}\mathbf{W}}{\mathbf{Z}\mathbf{W}^\top\mathbf{W}};$$

$$3. \mathbf{W} \leftarrow \mathbf{W} \odot \frac{(\mathbf{X}^\top\mathbf{Z} + \lambda\mathbf{M}\mathbf{Q})}{\mathbf{W}(\mathbf{Z}^\top\mathbf{Z} + \lambda\mathbf{Q}^\top\mathbf{Q})};$$

$$4. \mathbf{Q} \leftarrow \mathbf{Q} \odot \frac{\mathbf{M}^\top\mathbf{W}}{\mathbf{Q}\mathbf{W}^\top\mathbf{W}};$$

jusqu'à convergence

5. Normaliser \mathbf{Z} de sorte à obtenir des vecteurs unitaires.

\odot désigne le produit d'Hadamard.

Expériences

Données

TABLE – Description des jeux de données, # indique la cardinalité

Jeux de données	Caractéristiques					
	#Documents	#Termes	#Clusters	Densité $_X$ (%)	Équilibre	Densité $_M$ (%)
CSTR	475	1000	4	3.40	0.399	13.80
CLASSIC4	7095	5896	4	0.59	0.323	6.21
RCV1	6387	16921	4	0.25	0.080	3.07
NG5	4905	10167	5	0.92	0.943	22.04
NG20	18846	14390	20	0.59	0.628	25.06

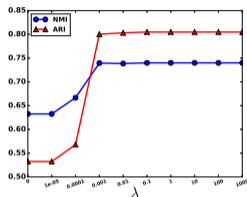
Transformations

- Pondération Term-Frequency Inverse Document Frequency (TF-IDF)
- Normalisation L_2

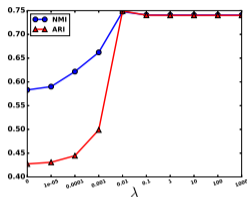
Plan d'expérience

- Choix de la fonction d'erreur
- Choix de λ
- Construction de M

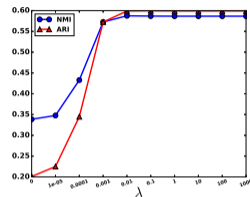
$\lambda = 0$ puis variant entre 10^{-5} et 10^3



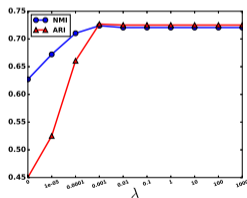
CSTR



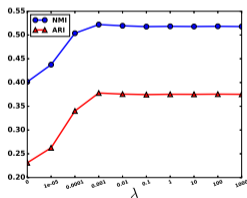
CLASSIC4



RCV1



NG5



NG20

- NMI : Normalized Mutual Information STREHL et GHOSH 2002
- ARI : Adjusted Rand Index HUBERT et ARABIE 1985

Résultats

TABLE – Moyenne des résultats sur 30 essais.

Données	Critères	NMF	ONMF	PNMF	GNMF	SNMF
CSTR	NMI	0.65±0.01	0.65±0.05	0.66±0.01	0.57±0.08	0.75±0.01
	ARI	0.54±0.01	0.56±0.04	0.56±0.01	0.53±0.11	0.80±0.01
CLASSIC4	NMI	0.51±0.09	0.55±0.09	0.59±0.05	0.65±0.04	0.72±0.06
	ARI	0.36±0.10	0.39±0.09	0.44±0.01	0.49±0.05	0.70±0.09
RCV1	NMI	0.39±0.03	0.49±0.002	0.46±0.001	0.48±0.04	0.56±0.01
	ARI	0.29±0.02	0.39±0.004	0.37±0.001	0.39±0.03	0.57±0.01
NG5	NMI	0.65±0.05	0.65±0.04	0.65±0.05	0.63±0.07	0.72±0.04
	ARI	0.48±0.09	0.48±0.08	0.47±0.09	0.62±0.09	0.70±0.06
NG20	NMI	0.43±0.01	0.44±0.02	0.45±0.02	0.52±0.01	0.53±0.01
	ARI	0.24±0.01	0.22±0.02	0.24±0.02	0.35±0.05	0.37±0.01

- NMF LEE et SEUNG 2001
- ONMF : Othogonal NMF DING, LI et PENG 2008; YOO et CHOI 2008
- PNMf : Projective NMF YUAN, YANG et OJA 2009; YANG et OJA 2010
- GNMF : Graph regularized NMF CAI et al. 2011

Avec d'autres mesures d'erreur

Rappel du modèle

$$\mathcal{F}(\mathbf{Z}, \mathbf{W}, \mathbf{Q}) = \underbrace{D_1(\mathbf{X}, \mathbf{Z}\mathbf{W}^\top)}_{\text{NMF}} + \lambda \underbrace{D_2(\mathbf{M}, \mathbf{W}\mathbf{Q}^\top)}_{\text{word embedding}}$$

Avec $D_1 = D_2 \iff$ la divergence de KullBack-Leibler :

$$\mathcal{F}(\mathbf{Z}, \mathbf{W}, \mathbf{Q}) = \sum_{i,j}^{n,d} x_{ij} \log \frac{x_{ij}}{(\mathbf{Z}\mathbf{W}^\top)_{ij}} - x_{ij} + (\mathbf{Z}\mathbf{W})_{ij} + \lambda \sum_{j,j'}^{d,d'} m_{jj'} \log \frac{m_{jj'}}{(\mathbf{W}\mathbf{Q}^\top)_{jj'}} - m_{jj'} + (\mathbf{W}\mathbf{Q}^\top)_{jj'} \quad (2)$$

Algorithme pour SNMF-KL

Algorithm 2 Kullback-Leibler Semantic-NMF (SNMF-KL).

Entrées : $\mathbf{X} \in \mathbb{R}_+^{n \times d}$, $\mathbf{M} \in \mathbb{R}_+^{d \times d'}$, λ et g .

Sorties : $\mathbf{Z} \in \mathbb{R}_+^{n \times g}$, $\mathbf{W} \in \mathbb{R}_+^{d \times g}$ et $\mathbf{Q} \in \mathbb{R}_+^{d' \times g}$.

Étapes :

1. Initialisation : $\mathbf{Z} \leftarrow \mathbf{Z}^{(0)}$; $\mathbf{W} \leftarrow \mathbf{W}^{(0)}$ et $\mathbf{Q} \leftarrow \mathbf{Q}^{(0)}$;

répéter

$$2. \mathbf{Z} \leftarrow \mathbf{Z} \odot \frac{\mathbf{X}}{\mathbf{Z}\mathbf{W}^\top \mathbf{W}} \Big/ \sum_j \mathbf{W}_{jk};$$

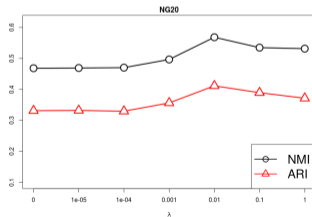
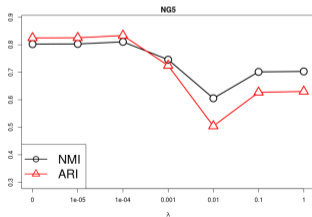
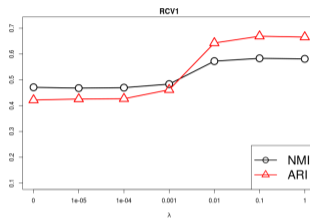
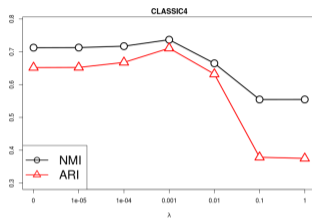
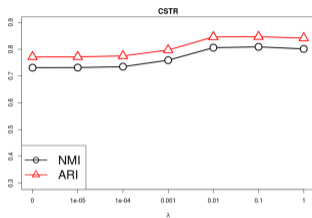
$$3. \mathbf{W} \leftarrow \mathbf{W} \odot \left(\frac{\mathbf{X}^\top \mathbf{Z}}{\mathbf{W}\mathbf{Z}^\top \mathbf{Z}} + \lambda \frac{\mathbf{M}}{\mathbf{W}\mathbf{Q}^\top \mathbf{Q}} \right) \Big/ \sum_i \mathbf{Z}_{ik} + \lambda \sum_{j'} \mathbf{Q}_{j'k};$$

$$4. \mathbf{Q} \leftarrow \frac{\mathbf{Q} \odot \frac{\mathbf{M}^\top}{\mathbf{Q}\mathbf{W}^\top \mathbf{W}}}{\sum_j \mathbf{W}_{jk}}$$

jusqu'à convergence

5. Normaliser \mathbf{Z} de sorte à obtenir des vecteurs unitaires.

$\lambda = 0$ puis variant entre 10^{-5} et 10^0 pour SNMF-KL



- NMI : Normalized Mutual Information STREHL et GHOSH 2002
- ARI : Adjusted Rand Index HUBERT et ARABIE 1985

Résultats pour la SNMF-KL

TABLE – Moyenne des résultats sur 30 essais.

Datasets	Metrics	NMF-KL		SNMF ($\lambda = 1$)		SNMF-KL ($\lambda = 10^{-3}$)	
CSTR	NMI	0.73±0.03	(0.76)	0.75±0.01	(0.76)	0.76±0.03	(0.81)
	ARI	0.77±0.04	(0.81)	0.80±0.01	(0.81)	0.80±0.04	(0.84)
CLASSIC4	NMI	0.71±0.02	(0.72)	0.72±0.06	(0.74)	0.74±0.01	(0.75)
	ARI	0.65±0.06	(0.65)	0.70±0.09	(0.74)	0.71±0.02	(0.73)
RCV1	NMI	0.47±0.02	(0.47)	0.56±0.01	(0.59)	0.48±0.03	(0.49)
	ARI	0.42±0.02	(0.43)	0.57±0.01	(0.60)	0.46±0.04	(0.44)
NG5	NMI	0.80±0.03	(0.83)	0.72±0.04	(0.72)	0.75±0.07	(0.72)
	ARI	0.82±0.04	(0.85)	0.70±0.06	(0.73)	0.72±0.1	(0.67)
NG20	NMI	0.47±0.01	(0.48)	0.53±0.01	(0.52)	0.50±0.02	(0.51)
	ARI	0.33±0.02	(0.34)	0.37±0.01	(0.37)	0.36±0.03	(0.38)

Résultats pour la SNMF-KL

TABLE – Moyenne des résultats sur 30 essais.

Datasets	Metrics	NMF-KL		SNMF ($\lambda = 1$)	
CSTR	NMI	0.73±0.03	(0.76)	0.75±0.01	(0.76)
	ARI	0.77±0.04	(0.81)	0.80±0.01	(0.81)
CLASSIC4	NMI	0.71±0.02	(0.72)	0.72±0.06	(0.74)
	ARI	0.65±0.06	(0.65)	0.70±0.09	(0.74)
RCV1	NMI	0.47±0.02	(0.47)	0.56±0.01	(0.59)
	ARI	0.42±0.02	(0.43)	0.57±0.01	(0.60)
NG5	NMI	0.80±0.03	(0.83)	0.72±0.04	(0.72)
	ARI	0.82±0.04	(0.85)	0.70±0.06	(0.73)
NG20	NMI	0.47±0.01	(0.48)	0.53±0.01	(0.52)
	ARI	0.33±0.02	(0.34)	0.37±0.01	(0.37)

En construisant M différemment

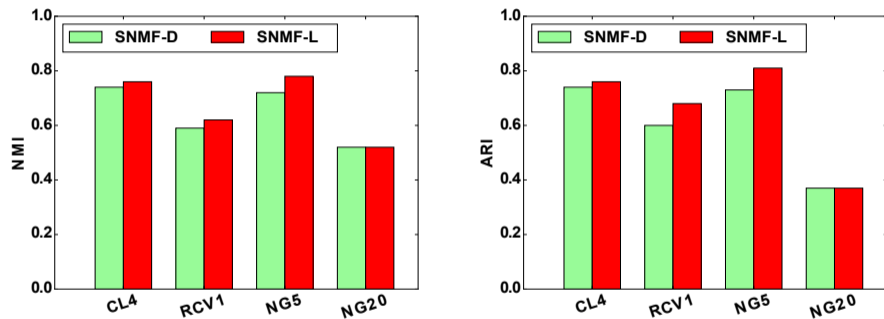


FIGURE – Gauche : comparaison d’NMI entre SNMF-D et SNMF-L. Droite : comparaison d’ARI entre SNMF-D et SNMF-L.

Conclusion

- Meilleure prise en compte des relations sémantiques existantes entre les différents termes d'un corpus
- Des résultats de classification significativement meilleurs
- Flexibilité en fonction des différentes ressources pouvant être utilisées pour enrichir la classification des éléments du corpus de X







Ouvertures

- Utilisation de différentes bases données telles que Wikipedia ou Google pour la construction de M





Fin

Merci de votre attention !

Références I

-  AILEM, Melissa, Aghiles SALAH et Mohamed NADIF (2017). “Non-negative Matrix Factorization Meets Word Embedding”. In : *ACM SIGIR*, p. 1081–1084.
-  CAI, Deng et al. (2011). “Graph regularized nonnegative matrix factorization for data representation”. In : *IEEE TPAMI* 33.8, p. 1548–1560.
-  DING, Chris, Tao LI et Wei PENG (2008). “On the equivalence between non-negative matrix factorization and probabilistic latent semantic indexing”. In : *Computational Statistics & Data Analysis* 52.8, p. 3913–3927.
-  HUBERT, Lawrence et Phipps ARABIE (1985). “Comparing partitions”. In : *Journal of classification* 2.1, p. 193–218.
-  LEE, Daniel D et H Sebastian SEUNG (1999). “Learning the parts of objects by non-negative matrix factorization”. In : *Nature* 401.6755, p. 788–791.
-  – (2001). “Algorithms for non-negative matrix factorization”. In : *Advances in neural information processing systems*, p. 556–562.

Références II

-  STREHL, Alexander et Joydeep GHOSH (2002). “Cluster ensembles—a knowledge reuse framework for combining multiple partitions”. In : *Journal of machine learning research* 3.Dec, p. 583–617.
-  YANG, Zhirong et Erkki OJA (2010). “Linear and nonlinear projective nonnegative matrix factorization”. In : *IEEE Transactions on Neural Networks* 21.5, p. 734–749.
-  YOO, Jiho et Seungjin CHOI (2008). “Orthogonal nonnegative matrix factorization : Multiplicative updates on Stiefel manifolds”. In : *IDEAL*, p. 140–147.
-  YUAN, Zhijian, Zhirong YANG et Erkki OJA (2009). “Projective nonnegative matrix factorization : Sparseness, orthogonality, and clustering”. In : *Neural Process. Lett.*