Méthodes de factorisation profonde pour la séparation de profils de pollution

Matthieu PUIGT Gilles DELMAIRE Gilles ROUSSEL

Univ. Littoral Côte d'Opale, EA 4491 – LISIC, F-62228 Calais, France

firstname.LASTNAME@univ-littoral.fr

Avant-propos : la factorisation matricielle (1)

Vous savez tous résoudre un système d'équations

$$\left\{\begin{array}{lcl} 2 \cdot s_1 + 3 \cdot s_2 & = & 5 \\ 3 \cdot s_1 - 2 \cdot s_2 & = & 1 \end{array}\right. \tag{1}$$

Si on note l'équation sous forme matricielle avec A, $\underline{s}$ et $\underline{x}$ la matrice et les vecteurs :

$$A = \begin{bmatrix} 2 & 3 \\ 3 & -2 \end{bmatrix}, \underline{s} = [s_1, s_2]^T \text{ et } \underline{x} = [5, 1]^T$$

Alors l'Eq. (1) devient

$$\underline{x} = A \cdot \underline{s}$$

dont la solution est :

$$\underline{s} = {\color{red}A^{-1}} \cdot \underline{x} = [1, 1]^T$$

Trouver $\underline{s}$ en fonction de $\underline{x}$ s'appelle un **problème inverse** car on doit inverser l'opérateur A.

Avant-propos : la factorisation matricielle (1)

Vous savez tous résoudre un système d'équations

$$\begin{cases} 2 \cdot s_1 + 3 \cdot s_2 \quad + \ldots + 7 \cdot s_5 = \quad 5 \\ 3 \cdot s_1 - 2 \cdot s_2 \quad + \ldots + 2 \cdot s_5 = \quad 1 \end{cases} \tag{1}$$

Si on note l'équation sous forme matricielle avec A, $\underline{s}$ et $\underline{x}$ la matrice et les vecteurs :

$$A = \begin{bmatrix} 2 & 3 & \ldots & 7 \\ 3 & -2 & \ldots & 2 \end{bmatrix}, \; \underline{s} = [s_1, s_2]^T \text{ et } \underline{x} = [5, 1]^T$$

Alors l'Eq. (1) devient

$$\underline{x} = A \cdot \underline{s}$$

dont la solution est :

$$\underline{s} = ???$$

Trouver $\underline{s}$ en fonction de $\underline{x}$ s'appelle un **problème inverse** car on doit inverser l'opérateur A.

Comment résoudre ce type de problème si ???

- Il y a plus d'inconnues que d'équations (problème inverse **mal posé**)

Avant-propos : la factorisation matricielle (1)

Vous savez tous résoudre un système d'équations

$$\left\{ \begin{array}{lcl} ? \cdot s_1 + ? \cdot s_2 & = & 5 \\ ? \cdot s_1 + ? \cdot s_2 & = & 1 \end{array} \right. \quad (1)$$

Si on note l'équation sous forme matricielle avec A, $\underline{s}$ et $\underline{x}$ la matrice et les vecteurs :

$$A = \begin{bmatrix} ? & ? \\ ? & ? \end{bmatrix}, \ \underline{s} = [s_1, s_2]^T \ \text{et} \ \underline{x} = [5, 1]^T$$

Alors l'Eq. (1) devient

$$\underline{x} = A \cdot \underline{s}$$

dont la solution est :

$$\underline{s} = A^{-1} \cdot \underline{x} = ?$$

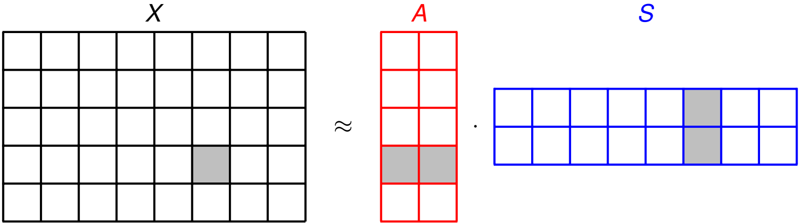
Trouver $\underline{s}$ en fonction de $\underline{x}$ s'appelle un **problème inverse** car on doit inverser l'opérateur A.

Comment résoudre ce type de problème si ???

- Il y a plus d'inconnues que d'équations (problème inverse **mal posé**)
- Si on ne connaît pas l'opérateur A (approche **aveugle** de déconvolution, égalisation, localisation, séparation de sources)

Avant-propos : la factorisation matricielle (2)

Enormément de problèmes, en *machine learning* et en traitement du signal, peuvent s'écrire sous forme de systèmes d'équations $X \approx A \cdot S$:



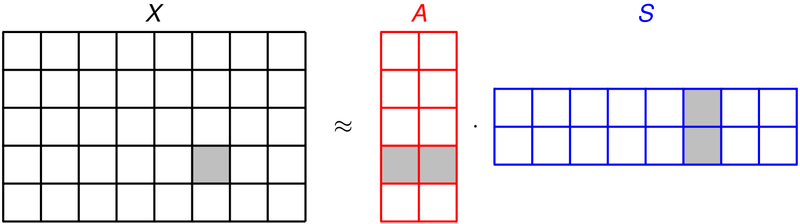
Approximation de matrices de faible rang

X matrice bruitée ou à trous, A matrice de poids, S matrice de variables latentes

- *Topic modeling*
- Filtrage collaboratif (par ex: problème Netflix)
- Etalonnage de capteurs mobiles (Thèse de Clément Dorffer)

Avant-propos : la factorisation matricielle (2)

Enormément de problèmes, en *machine learning* et en traitement du signal, peuvent s'écrire sous forme de systèmes d'équations $X \approx A \cdot S$:



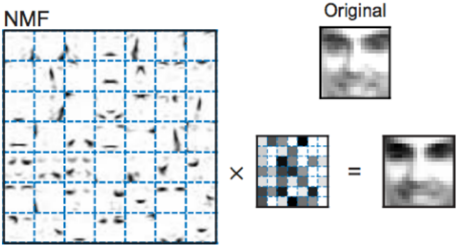
localisation/séparation de sources

X matrice de données observées, A matrice de mélange, S matrice des sources

- Si A estimée (et géométrie du réseau de capteurs connue)
- ➪ estimation de mélanges (localisation de souces)
- Si S estimée (et géométrie du réseau de capteurs connue)
- ➪ séparation de sources (traitement d'antennes)

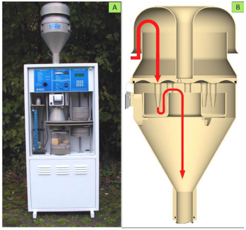
Avant-propos : la factorisation matricielle (3)

- Dans beaucoup de problèmes, les matrices dans $X \approx A \cdot S$ sont non-négatives
- ➯ Contraintes de non-négativité sur A et/ou S amène une meilleure interprétabilité
- ➯ Non-negative Matrix Factorization (NMF)



NMF appliquée à des visages (source : Lee & Seung, 1999)

Séparation de sources chimiques

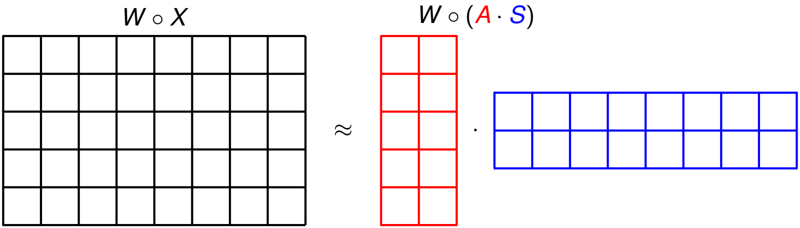


- Acquisition de particules fines
- n échantillons analysés par les chimistes
- Données observées rangées dans une matrice (un tableau) X de taille $n \times m$ représentant les concentrations de m espèces chimiques (en ng/m^3)
- Données observées sont des mélanges de "profils"

$$X \simeq A \cdot S$$

- ➥ A est la matrice $n \times p$ de contribution (ng/m^3)
- ➥ S est la matrice $p \times m$ de profils (proportions relatives d'espèces chimiques) de sources (ng/ng)

Analyse de l’existant



NMF informée

- NMF classique ne fournit pas de solution convenable
- Information complémentaire :
  - Mesure de confiance dans les données X (opérateur W)
  - Connaissance expertes (valeurs figées, bornes, critères de comparaison des profils)

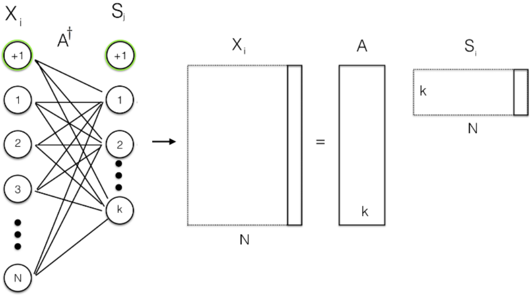
➪ NMF informée (Delmaire *et al.*, 2008–2017)

Questions ouvertes

- Variabilité des sources au cours des saisons ?
- Influence des conditions météo ?

Factorisation matricielle et réseau de neuronnes (1)

- la NMF a été aussi très utilisée en **apprentissage supervisé** (*machine learning*)
  - On apprend S d'une factorisation $X_1 = A \cdot S$ (son ou image spécifique)
  - On utilise cette matrice dans une autre factorisation : $X_2 = G \cdot S + F \cdot H$
- En apprentissage supervisé, on utilise depuis longtemps les réseaux de neuronnes
- Quel est le lien entre une factorisation matricielle et un réseau de neuronnes ?



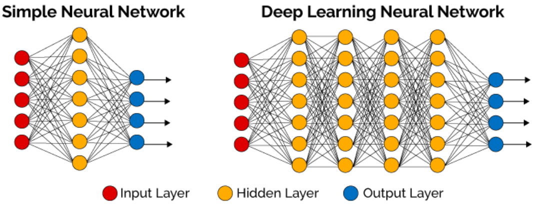
Source: Flener & Hunter, 2015

Factorisation matricielle et réseau de neuronnes (2)

La révolution du *deep learning*

- Jusqu'aux années 2005, réseaux de neuronnes utilisés avec peu de couches (coût calculatoire)
- Machines plus performantes ➪ Plus de couches

➪ Des performances inégalées (au prix d'un apprentissage sur une énorme quantité de données)



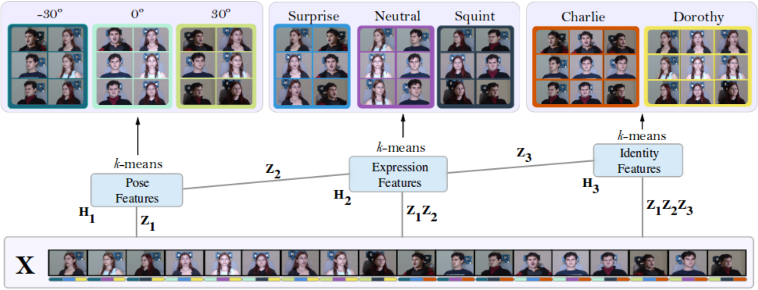
Source: `https://goo.gl/kwoJ18`

Deep NMF

- Si une couche de réseau de neuronne $\simeq$ 1 NMF
- Pourquoi ne pas faire des NMF successives

➪ Deep (semi-)NMF (cf littérature avec le sujet du projet)

Deep NMF pour la séparation de sources chimiques ?



Source: Trigeorgis *et al.*, 2014

- Deep (semi-)NMF extraie des informations structurantes à chaque couche de factorisation (similarités avec classification hiérarchique)
- Est-ce que ce type de factorisation profonde peut apporter un plus pour la séparation de sources chimiques (conditions climatiques, phénomènes saisoniers, etc) ?
- Eventuellement, peut-on proposer des approches informées de *Deep NMF* ?

Déroulé du projet

Objectifs

1. Lire et comprendre la littérature
2. Commencer à prendre en main/programmer des méthodes de Deep NMF – dont (Trigeorgis *et al.*, 2014)
   `https://github.com/trigeorgis/Deep-Semi-NMF`
3. Rajouter l'influence de la pondération de W dans ces approches
4. Tester ces approches sur des données de pollution chimique

Y parvenir

- Première réunion formelle où vous présenterez à l'équipe ce que vous avez lu/compris
- Réunions informelles pour discuter de votre avancement
- Programmation de préférence pour Matlab, Python ou C (communautés)
- Stockage de vos codes sur le service Git de l'ULCO

Questions ?