Sur la normalisation de la matrice Laplacienne en partitionnement spectral

Julien Ah-Pine
(julien.ah-pine@univ-lyon2.fr)

Université Lyon 2
Laboratoire ERIC

Lyon, le 28 juin 2017
SFC 2017

Sommaire

Rappel du Sommaire

Partitionnement spectral (PS) en quelques mots

- Méthodes de clustering développées depuis les années 2000.
- Prise en compte de la **géométrie intrinsèque** des données (espaces courbées/variétés).
- Les notions de **graphes de similarités et de voisinage** et de **matrice laplacienne** sont importantes.
- Méthodologie : **décomposition spectrale** de la matrice Laplacienne, plongement des données dans un **espace euclidien**, et utilisation des k-**means** pour le partitionnement.
- Liens avec les problèmes et méthodes de partitionnement de graphe.
- Article d'introduction de référence : [Von Luxburg, 2007].

Graphe de similarités

- La modélisation employée est celle de **graphe non-orienté pondéré**.
- Soit $G = (\mathbb{V}, \mathbb{E})$ un graphe non-orienté, sans boucle.
- $\mathbb{V} = \{v_1, \ldots, v_n\}$ est l'ensemble des sommets : objets à classifier.
- $\mathbb{E}$ est l'ensemble des arêtes : paires d'objets qui sont **similaires**.
- Les arêtes sont pondérées et si $(v_i, v_j) \in \mathbb{E}$, alors $l(v_i, v_j) > 0$ est la **valuation de la similarité**.
- G est représenté par une **matrice d'adjacence pondérée** notée $\mathbf{W} = (w_{ij})_{i,j=1,\ldots,n}$, définie comme suit :

$$w_{ij} = \begin{cases} l(v_i, v_j) & \text{si } (v_i, v_j) \in \mathbb{E} \\ 0 & \text{sinon} \end{cases}$$

Graphe de voisinage

- L'ensemble des arêtes $\mathbb{E}$ encode des **relations de voisinage**.
- Plusieurs façons de construire $\mathbb{E}$ (et donc $\mathbf{W}$) :
  - **Graphe écrêté selon un seuil** $\theta \geq 0$:

$$(v_i, v_j) \in \mathbb{E} \Leftrightarrow I(v_i, v_j) \geq \theta$$

  - **Graphe des** k**-plus proches voisins** :

$$(v_i, v_j) \in \mathbb{E} \Leftrightarrow (v_i \subset \mathbb{NN}_k(v_j) \vee v_j \subset \mathbb{NN}_k(v_i))$$

    où $\mathbb{NN}_k(v_i)$ est l'ensemble des k plus proches voisins de v_i.

Matrice laplacienne et sa normalisation classique

- Notons $\mathbf{D} = (d_{ij})_{i,j=1,\dots,n}$ la **matrice des degrés** et définie par :

$$d_{ij} = \begin{cases} d_i & \text{si } i = j \\ 0 & \text{sinon} \end{cases}$$

  où $d_i = \sum_{j=1}^{n} w_{ij}\ , \forall i = 1, \dots, n$.
- La **matrice Laplacienne** de G notée $\mathbf{L}$ est définie comme suit :

$$\mathbf{L} = \mathbf{D} - \mathbf{W}$$

- La **normalisation par division symétrique** de $\mathbf{L}$ notée $\mathbf{L}_{sym}$ est donnée par :

$$\begin{aligned} \mathbf{L}_{sym} &= \mathbf{D}^{-1/2}\mathbf{L}\mathbf{D}^{-1/2} \\ &= \mathbf{I} - \mathbf{D}^{-1/2}\mathbf{W}\mathbf{D}^{-1/2} \end{aligned}$$

  où $\mathbf{I}$ est la matrice identité d'ordre n.

Propriétés de la matrice laplacienne normalisée

Propriété.

Soit $\mathbf{L}_{sym}$ *la matrice laplacienne* **normalisée** *d'un graphe* G *:*

- *Pour tout vecteur* $\mathbf{f} \in \mathbb{R}^n$ *:*

$$\mathbf{f}^\top \mathbf{L}_{sym}\mathbf{f} = \frac{1}{2}\sum_{i,j=1}^{n} w_{ij}\left(\frac{f_i}{\sqrt{d_i}} - \frac{f_j}{\sqrt{d_j}}\right)^2$$

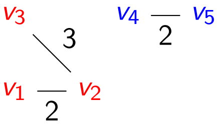
- $\mathbf{L}_{sym}$ *est symétrique et sdp et possède* n *valeurs propres réelles, non-négatives :*

$$0 = \lambda_1 \leq \lambda_2 \leq \ldots \leq \lambda_n$$

- *L'ordre de multiplicité* k *de la* **valeur propre 0** *est le nombre de composantes connexes de* G *notées* $C_1, \ldots, C_k$. *Le sous-espace propre associé à la valeur propre 0 est engendré par les vecteurs* $\mathbf{D}^{1/2}\mathbf{1}_{C_1}, \ldots, \mathbf{D}^{1/2}\mathbf{1}_{C_k}$ *où* $\mathbf{1}_{C_l}$ *est le vecteur indicateur de* C_l.

Illustration

- $\mathbb{V} = \{v_1, v_2, v_3, v_4, v_5\}$
- $\mathbb{E} = \{\underbrace{(v_1, v_2)}_{2}, \underbrace{(v_2, v_3)}_{3}, \underbrace{(v_4, v_5)}_{2}\}$



$$\rightarrow \mathbf{W} = \begin{pmatrix} 0 & 2 & 0 & 0 & 0 \\ 2 & 0 & 3 & 0 & 0 \\ 0 & 3 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 2 \\ 0 & 0 & 0 & 2 & 0 \end{pmatrix} \qquad \rightarrow \mathbf{D} = \begin{pmatrix} 2 & 0 & 0 & 0 & 0 \\ 0 & 5 & 0 & 0 & 0 \\ 0 & 0 & 3 & 0 & 0 \\ 0 & 0 & 0 & 2 & 0 \\ 0 & 0 & 0 & 0 & 2 \end{pmatrix}$$

$$\rightarrow \mathbf{L} = \begin{pmatrix} 2 & -2 & 0 & 0 & 0 \\ -2 & 5 & -3 & 0 & 0 \\ 0 & -3 & 3 & 0 & 0 \\ 0 & 0 & 0 & 2 & -2 \\ 0 & 0 & 0 & -2 & 2 \end{pmatrix} \qquad \rightarrow \mathbf{L}_{sym} = \begin{pmatrix} 1 & -.63 & 0 & 0 & 0 \\ -.63 & 1 & -.77 & 0 & 0 \\ 0 & -.77 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & -1 \\ 0 & 0 & 0 & -1 & 1 \end{pmatrix}$$

$\rightarrow$ Spectre de $\mathbf{L}_{sym}$: $\{2,2,1,0,0\}$

$$\rightarrow \underbrace{\mathbf{D}^{1/2}\mathbf{1}_{C_1}}_{\mathbf{v}_0^1} = \begin{pmatrix} 1.41 \\ 2.24 \\ 1.73 \\ 0 \\ 0 \end{pmatrix} \text{ et } \underbrace{\mathbf{D}^{1/2}\mathbf{1}_{C_2}}_{\mathbf{v}_0^2} = \begin{pmatrix} 0 \\ 0 \\ 0 \\ 1.41 \\ 1.41 \end{pmatrix}.$$

Pseudo-code des méthodes de partitionnement spectral

Input : I (similarité), méthode de voisinage, k (nb de classes)
Construire le graphe de voisinage $\mathbf{W}$ selon la méthode choisie
Construire la matrice laplacienne normalisée $\mathbf{L}_{sym}$
Calculer $\mathbf{f}^1, \ldots, \mathbf{f}^k$, les k premiers vecteurs propres de $\mathbf{L}_{sym}$
Construire $\mathbf{F} = \begin{pmatrix} \mathbf{f}^1 & \ldots & \mathbf{f}^k \end{pmatrix} \in \mathbb{R}^{n \times k}$
Normer les vecteurs lignes de $\mathbf{F}$
Utiliser les k-means pour partitionner les n lignes de $\mathbf{F}$
Ouput : Partition en k classes

- Remarque : on peut utiliser $\mathbf{L}$ à la place de $\mathbf{L}_{sym}$ mais la dernière a de meilleures propriétés de convergence asymptotique que la première [Von Luxburg et al., 2008].

Rappel du Sommaire

Normalisation par division symétrique et produit de Hadamard

- **Proposition** : extension de la normalisation par division symétrique.
- Reprenons $\mathbf{L}_{sym}$ ainsi que sa forme quadratique associée :

$$\mathbf{L}_{sym} = \mathbf{I} - \mathbf{D}^{-1/2}\mathbf{W}\mathbf{D}^{-1/2} \Rightarrow \mathbf{f}^\top \mathbf{L}_{sym}\mathbf{f} = \sum_{i=1}^{n} f_i^2 - \sum_{i,j=1}^{n} \frac{w_{ij}}{\sqrt{didj}} f_i f_j$$

- Reformulation de $\mathbf{L}_{sym}$ en utilisant le **produit de Hadamard (ou de Schur)** noté $\circ$ (produit terme à terme entre deux matrices).
- En prenant $\mathbf{N} = (n_{ij})_{i,j=1,\dots,n}$ de terme général :

$$n_{ij} = \frac{1}{\sqrt{didj}}$$

- On montre facilement que :

$$\boxed{\mathbf{L}_{sym} = \mathbf{L} \circ \mathbf{N}}$$

Normalisation d'ordre t

- Etant donné le vecteur de degrés $\mathbf{d} = (d_1, \ldots, d_n)$, on introduit la famille paramétrique de matrices $\mathbf{N}^t = (n_{ij}^t)_{i,j=1,\ldots,n}$ de terme général :

$$n_{ij}^t = \frac{1}{\left(\frac{1}{2}(d_i^t + d_j^t)\right)^{\frac{1}{t}}}$$

où le paramètre t est un réel strictement positif.

Définition.

La **famille de matrice Laplacienne normalisée d'ordre** t*, notée* $\mathbf{L}_{sym}^t$*, est défine pour tout* $t > 0$ *par :*

$$\boxed{\begin{aligned} \mathbf{L}_{sym}^t &= \mathbf{L} \circ \mathbf{N}^t \\ &= \mathbf{I} - \mathbf{W} \circ \mathbf{N}^t \end{aligned}}$$

Normalisation d'ordre t (suite)

- Posons :

$$\mathcal{M}^t(d_i, d_j) = \left(\frac{1}{2}(d_i^t + d_j^t)\right)^{\frac{1}{t}}$$

- On a :

$$n_{ij}^t = \frac{1}{\mathcal{M}^t(d_i, d_j)}, \forall i, j = 1, \ldots, n$$

- $\mathcal{M}^t$ est la **moyenne généralisée (puissance) d'ordre** $t > 0$ avec les propriétés remarquables suivantes :
  - Moyenne géométrique : $\lim_{t \to 0} \mathcal{M}^t(d_i, d_j) = \sqrt{didj}$,
  - Moyenne arithmétique : $\mathcal{M}^1(d_i, d_j) = \frac{1}{2}(di + dj)$,
  - Maximum : $\lim_{t \to \infty} \mathcal{M}^t(d_i, d_j) = \max(di, dj)$.
- Clairement on a :

$$\boxed{\lim_{t \to 0} \underbrace{\mathbf{L} \circ \mathbf{N}^t}_{\mathbf{L}_{sym}^t} = \mathbf{L}_{sym}}$$

Interprétations

- Forme quadratique associée à $\mathbf{L}^t_{sym}$, $\forall \mathbf{f} \in \mathbb{R}^n$:

$$\boxed{\mathbf{f}^\top \mathbf{L}^t_{sym}\mathbf{f} = \sum_{i=1}^{n} f_i^2 - \sum_{i,j=1}^{n} \frac{w_{ij}}{\left(\frac{1}{2}(d_i^t + d_j^t)\right)^{\frac{1}{t}}} f_i f_j}$$

- Pour minimiser $\mathbf{f}^\top \mathbf{L}^t_{sym}\mathbf{f}$, il faut que $f_i f_j > 0$:
  - "$f_i f_j > 0$ si w_{ij} grand".
  - "$f_i f_j > 0$ si $\left(\frac{1}{2}(d_i^t + d_j^t)\right)^{\frac{1}{t}}$ petit".
- Propriété : $\forall t > 0, \begin{cases} \sqrt{d_i d_j} \leq \left(\frac{1}{2}(d_i^t + d_j^t)\right)^{\frac{1}{t}} \\ \sqrt{d_i d_j} = \left(\frac{1}{2}(d_i^t + d_j^t)\right)^{\frac{1}{t}} \quad \text{ssi } d_i = d_j \end{cases}$.

▷ "$t > 0$ **pénalise** davantage la différence entre d_i et d_j que $t \to 0$".

▷ v_i et v_j ont une forte probabilité d'être dans la même classe ($f_i f_j > 0$) si leur **similarité** est **forte ET** si leurs **dégrés** sont **similaires**.

Propriété de métricité

Propriété.

Pour tout $t > 0$, $\mathbf{L}^t_{sym}$ est semi-définie positive.

Démonstration.

$\mathbf{L}$ est sdp. On montre que $\mathbf{N}^t$ est sdp pour tout $t > 0$ [Bhatia, 2006]. Ainsi, $\mathbf{L} \circ \mathbf{N}^t$ est sdp comme conséquence du théorème de Schur. $\square$

- Ce résulat permet de garantir le plongement des objets dans un espace Euclidien par décomposition spectrale de $\mathbf{L}^t_{sym}, t > 0$.
- En pratique, on peut donc utiliser le pseudo-code précédent avec cette nouvelle famille de matrice Laplacienne normalisée.

Rappel du Sommaire

Protocol expérimental

- Tests sur 5 jeux de données accessibles en ligne.
- Tests de 5 normalisations : $t \to 0$ (référence) contre $t = 1, 2, 5, 10$.
- Critère de qualité externe : indice de Rand corrigé (ARI).
- Pour chaque cas, 10 exécutions de k-means, puis test de significativité de Wilcoxon-Mann-Withney des mesures ARI entre $t \to 0$ et $t > 0$.
- Tous les résultats présentés (moyennes des ARI) sont statistiquement différents du résultat de référence.

Résultats

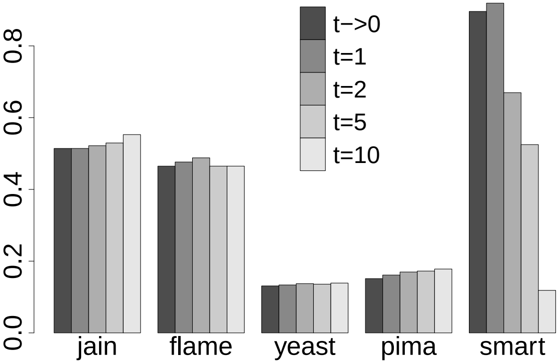


FIGURE – Résultats sur 5 jeux de données et 5 normalisation différentes.

Rappel du Sommaire

Conclusion et travaux futurs

- Contributions :
  - Généralisation de la normalisation par division symétrique.
  - "Proof of concept" sur quelques jeux de données.
- Travaux futurs :
  - Comment fixer le paramètre t ?
    (coefficient de clustering, ...),
  - Un paramètre t_{ij} pour chaque paire (v_i, v_j) ?
    (coefficient de clustering local, ...),
  - Partitionnement de graphes avec propriété sur les degrés.
    (réseaux invariants d'échelle, ...).

Merci de votre attention !
Des questions ? :-)

Bhatia, R. (2006).
Infinitely divisible matrices.
The American Mathematical Monthly, 113(3) :221–235.

Von Luxburg, U. (2007).
A tutorial on spectral clustering.
Statistics and computing, 17(4) :395–416.

Von Luxburg, U., Belkin, M., and Bousquet, O. (2008).
Consistency of spectral clustering.
The Annals of Statistics, pages 555–586.