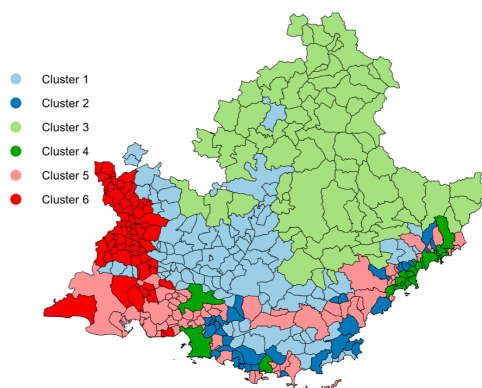


Co-clustering de données spatio-temporelles

Julien JACQUES

Lorsque l'on dispose d'un grand ensemble de données où à la fois le nombre d'observations (lignes) et le nombre de variables (colonnes) sont grands, le co-clustering permet de résumer les données en créant des clusters d'individus semblables mais également des clusters de variables ayant une distribution similaire.

Dans [1], un algorithme de co-clustering est proposé dans le cas où les variables observées sont des courbes multivariées. Le modèle probabiliste sur lequel repose cet algorithme suppose les individus indépendants les uns des autres. De même pour les variables.



L'objectif de ce TER est d'étendre ce modèle au cas où les individus sont spatialement dépendants et les variables temporellement dépendantes.

Le travail à réaliser dans ce TER est alors :

1. s'appropriier le modèle de co-clustering développer dans [1] ainsi que son implémentation dans le package `FunHDDC`,
2. comprendre comment il est possible de prendre en compte une dépendance temporelle ou spatiale [2, 3] dans un tel modèle,
3. en partant du package `FunHDDC`, créer un nouveau code qui permette de prendre en compte ces dépendances spatiales et temporelles,
4. tester le code sur données simulées et montrer l'avantage par rapport à la version initiale de `FunHDDC`,
5. mettre en application sur données réelles.

Dans le cas de résultats encourageants, une poursuite en stage au sein du laboratoire ERIC est envisageable.

Références

- [1] C. Bouveyron, J. Jacques, A. Schmutz, F. Simoes, S. Bottini (2021). Co-Clustering of Multivariate Functional Data for the Analysis of Air Pollution in the South of France. *Annals of Applied Statistics*, to appear. [<https://hal.archives-ouvertes.fr/hal-02862177>] [R package `FunHDDC`].
- [2] Cheam, A.S.M., Marbac, M. and McNicholas, P.D. (2017). Model-based clustering for spatio-temporal data on air quality monitoring. *Environmetrics*, 8 (3) [R package `SpaTimeClus`].
- [3] Vincent Vandewalle, Cristian Preda, Sophie Dabo. Clustering spatial functional data. *ERCIM 2018*, Dec 2018, Pise, Italy.