



#Statistiques

#Clustering

#DonnéesMixtes

#RéductionDeDimension

#ApprentissageNonSupervisé

Packages R

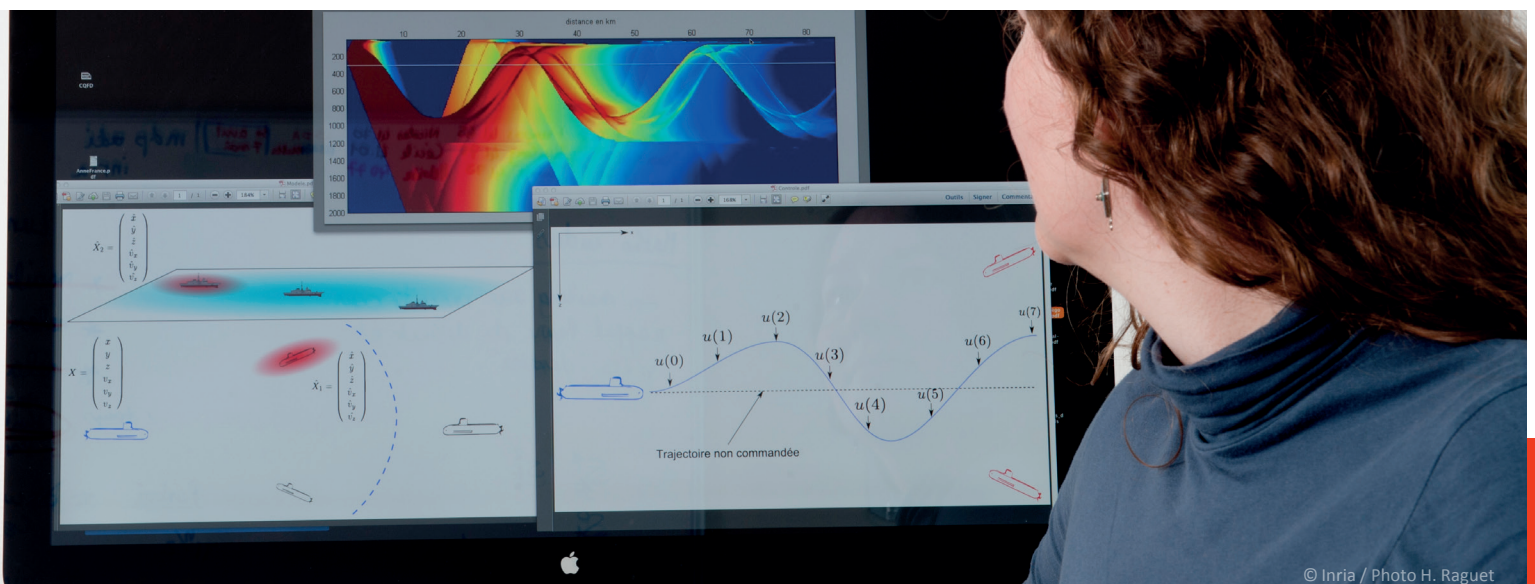
Faites parler vos données avec R

SOLUTION

L'analyse exploratoire des données s'articule autour de méthodes d'analyse des données multidimensionnelles (analyses factorielles, clustering par exemple) pouvant s'appliquer à des données qualitatives, quantitatives ou mixtes (mélange de quantitatif et de qualitatif).

AVANTAGES CONCURRENTIELS

- Fonctionne avec tous types de données - données mixtes (quantitatives et qualitatives) ou non
- Forte interprétabilité des résultats, nombreux résultats intermédiaires de calcul et paramètres consultables
- Affichage de mapping, cercle de corrélation, graphique en bâton, et autres représentations graphiques pour faciliter l'interprétation
- Permet de modéliser la structure des données afin d'en apprendre d'avantage sur les données
- Pour les méthodes non supervisées : méthodes extrêmement utiles quand on ne sait pas ce que l'on recherche





FICHE D'IDENTITÉ

- **Langage de programmation** : R
- **Licence** : GPL2/GPL3 (selon les *packages*)
- **Équipe-projet** : [Astral](#)
- **Compétences minimum requises** : R ou *Python*

CAS D'USAGES

Les méthodes de ces packages peuvent être appliquées aux différentes thématiques suivantes :

- **Santé** : génétique...
- **Environnement** : qualité de vie, météo...
- **Agronomie** : qualité nutritionnelle et sensorielle de la viande...
- **Marketing/vente** : segmentation clients, prévisions des stocks...
- **Industrie 4.0** : identification de mode de fonctionnement, détection d'anomalies...

FONCTIONNALITÉS GÉNÉRIQUES

- **PCAmixdata** : analyse en composantes principales (ACP) de données mixtes (quantitatives et/ou qualitatives), rotation, analyse factorielle multiple (AFM), réduction de dimension, recodage de données. Ce package est dédié à l'analyse factorielle de données mixtes. Il propose principalement une méthode d'ACP pour données mixtes (quantitatives et qualitatives) appelée PCAmix, une méthode de rotation orthogonale des composantes de PCAmix et une méthode d'AFM de données mixtes à l'intérieur des groupes.
- **ClustOfVar** : clustering de variable, réduction de dimension, recodage de données. Ce package propose principalement deux méthodes de clustering de variables qui fonctionnent pour des données mixtes (quantitatives et qualitatives) et une méthode de type bootstrap pour l'aide au choix du nombre de clusters. Chaque cluster est caractérisé par une variable synthétique quantitative et ClustOfVar peut être utilisé comme outil de réduction de dimension et de recodage de données en données quantitatives, au même titre que PCAmix.
- **ClustGeo** : clustering de données multivariées avec contraintes souples de contiguïté. Ce package est dédié au clustering de données multivariées, pour lesquelles des contraintes de contiguïté entre les observations sont connues (spatiales par exemple). Ces contraintes qui peuvent être géographiques sont intégrées dans une approche hiérarchique, un paramètre d'importance de ces contraintes sur le résultat du clustering pouvant être calibré.

EXPERTS

- **Marie Chavent** : professeure de l'université de Bordeaux, membre de l'équipe-projet Inria Astral
- **Hadrien Lorenzo** : docteur en mathématiques appliquées, membre de l'équipe-projet Inria Astral
- **Jérôme Saracco** : professeur des Universités, Institut Polytechnique de Bordeaux, membre de l'équipe-projet Inria Astral

POUR ALLER PLUS LOIN :

- <https://github.com/chavent/PCAmixdata>
- <https://github.com/chavent/ClustOfVar>
- <https://github.com/chavent/ClustGeo>