

Sujet de stage de Master

Détection de rupture multivariée et classification de courbes pour l'analyse de cinétique de métabolites

Mots - clés: séries temporelles multivariées ; clustering ; détection de rupture ; métabolomique

Contexte

Ce stage s'inscrit dans le cadre du projet de recherche [HepatoTwin](https://digitbio.hub.inrae.fr/rubriques-verticales2/nos-actions/projets-phares/projet-phare-hepato-twin-2024-2028) (<https://digitbio.hub.inrae.fr/rubriques-verticales2/nos-actions/projets-phares/projet-phare-hepato-twin-2024-2028>), dont nous sommes partenaires, en collaboration avec l'unité de recherche [TOXALIM](https://toxalim.toulouse.hub.inrae.fr/) (<https://toxalim.toulouse.hub.inrae.fr/>), laboratoire spécialisé en toxicologie alimentaire et coordinateur du projet. L'objectif principal de ce programme est d'étudier l'impact des contaminants alimentaires sur le métabolisme hépatique.

Dans ce contexte, nous disposons de séries temporelles portant sur plusieurs dizaines, voire centaines, de métabolites (petites molécules, produits ou substrats des réactions qui composent le métabolisme), mesurés en temps réel à partir de cultures de cellules hépatiques humaines. Ces cellules sont exposées à différents stress, simulant l'exposition à des contaminants alimentaires ou à des régimes alimentaires déséquilibrés, susceptibles de provoquer soit des ruptures brutales, soit des modifications plus progressives dans les dynamiques de certaines métabolites.

Sujet de stage

L'objectif du stage de recherche est de développer des méthodes d'apprentissage statistique pour détecter des perturbations métaboliques. Deux scénarios principaux seront étudiés :

1. un premier correspondant à une rupture nette,
2. un second caractérisé par une modification du schéma (pattern) dans les dynamiques de réponse.

Dans le premier scénario, les méthodes auront pour objectif de détecter une rupture « globale » en prenant en compte l'ensemble des métabolites (ou variables), et d'identifier des groupes de métabolites présentant des temps ou profils de réponse similaires. Cet objectif oriente le travail vers des approches multivariées et des méthodes de classification de courbes.

Le ou la stagiaire pourra s'appuyer sur l'expertise de ses encadrants :

- Guillem Rigaill, spécialiste des méthodes et algorithmes de détection de ruptures univariées et multivariées,
- Arthur Leroy, spécialiste des processus gaussiens, applicables notamment à la classification de courbes,
- Julien Chiquet, expert en apprentissage statistique, méthodes régularisées et analyses multivariées.

Les développements pourront être réalisés en Python et/ou R. Des données prototypiques sont déjà disponibles pour la mise au point des méthodes.

Références

1. Pishchagina, Rigaill and Runge. [*"Geometric-Based Pruning Rules for Change Point Detection in*](#)

- [Multiple Independent Time Series.](https://computo-journal.org/published-202406-pishchagina-change-point/) (https://computo-journal.org/published-202406-pishchagina-change-point/) Computo (2024) [Link to R/C package \(https://github.com/lpishchagina/GeomFPOP\)](https://github.com/lpishchagina/GeomFPOP)
2. Maidstone, Rigaiil, Hocking, Fearnhead. ["On optimal multiple changepoint algorithms for large data"](https://pmc.ncbi.nlm.nih.gov/articles/PMC7175693/) (https://pmc.ncbi.nlm.nih.gov/articles/PMC7175693/) Statistics and Computing (2016] [Link to R/C package \(https://cran.r-project.org/web/packages/fpop/index.html\)](https://cran.r-project.org/web/packages/fpop/index.html)
 3. Romano et al. ["Decafs: Detecting Changes in Autocorrelated and Fluctuating Signals"](https://www.tandfonline.com/doi/full/10.1080/01621459.2021.1909598) (https://www.tandfonline.com/doi/full/10.1080/01621459.2021.1909598) Journal of the American Statistical Association (2021) [Link to R/C package \(https://github.com/qtromano/DeCAFS\)](https://github.com/qtromano/DeCAFS)
 4. Leroy, Latouche, Guedj, Gey. ["Cluster-Specific Predictions with Multi-Task Gaussian Processes."](https://www.jmlr.org/papers/v24/20-1321.html) (https://www.jmlr.org/papers/v24/20-1321.html) Journal of Machine Learning Research (2023) [Link to R/C package \(https://arthurleroy.github.io/MagmaClustR/\)](https://arthurleroy.github.io/MagmaClustR/)

Conditions d'exercice

Le stage aura lieu dans l'UMR MIA Paris-Saclay, dans le Campus Agro Paris-Saclay, dans les conitions d'exercices proposé par INRAE (télétravail, gratification horaire de 4.35€, accès aux trasnports). Il sera encadré par Julien Chiquet [julien.chiquet@inrae.fr \(mailto:julien.chiquet@inrae.fr\)](mailto:julien.chiquet@inrae.fr) (DR INRAE), Guillem Rigaiil [guillem.rigaiil@inrae.fr \(mailto:guillem.rigaiil@inrae.fr\)](mailto:guillem.rigaiil@inrae.fr) (DR INRAE) et [arthur.leroy@inrae.fr \(mailto:arthur.leroy@inrae.fr\)](mailto:arthur.leroy@inrae.fr) Arthur Leroy (CR INRAE). Il bénéficiera de l'environnement et des données du projet Hepato'Twin.