CREST

Offre de stage de Master : utilisation des données issues d'accéléromètres pour la prédiction

ENSAI École nationale de la statistique et de l'analyse de l'information

Contexte L'étude des liens entre l'activité physique et certaines pathologies (obésité, hypertension, bléssures récurrentes chez le sportif) est un enjeu majeur de santé publique. Pour mesurer l'activité physique de différents sujets, l'utilisation d'un accéléromètre est de plus en plus courante. Cependant l'analyse des données issues de ces appareils pose des questions statistiques complexes : gestion du volume de données (les appareils mesurent l'activité chaque seconde pendant plusieurs jours), intégration de ces données dans un modèle statistique. En général, ces données complexes sont résumées par des temps passés à certains niveaux d'activité, ce qui les rend plus faciles à manipuler. Le but du stage est de construire une procédure statistique permettant d'étudier le lien entre une variable cible (une pathologie) et un ensemble de variables explicatives dont l'activité physique.

Description du stage Dans un premier temps, une méthode en deux étapes sera considérée. La permière étape estime des profils d'individus en fonction de leur activité physique en faisant un clustering des données d'accélérométrie [1, 2]. La seconde étape consiste à plugger la variable d'appartenance aux classes dans une régression semi-paramétrique multivariée et à tester sa significativité. Dans un second temps, on cherchera à effectuer ces deux étapes simultanément pour améliorer la qualité des estimateurs [3]. Les méthodes utilisées seront implémentées en R, illustrées sur des données simulées afin d'en valider la pertinence puis appliquées sur les données issues de l'étude *Physical Activity and Transit* [4].

Profil

- Étudiant de Master 2 mathématiques appliquées ou statistiques ou en 3ème année d'école d'ingénieurs (les candidatures d'étudiants en Master 1 seront considérées).
- Connaissance des méthodes d'estimation de densité, de régression et de clustering.
- La connaissance de C++ et/ou de R est un plus.

Conditions du stage

- Lieu : ENSAI au campus Ker Lann à Bruz.
- Durée : 6 mois.

Contact Les candidats devront envoyer un CV et une lettre de motivation à Marie Du Roy de Chaumaray (marie.du-roy-de-chaumaray@ensai.fr) et Matthieu Marbac (matthieu.marbac-lourdelle@ensai.fr).

Informations Le stage a reçu un financement de l'EUR DIGISPORT (Digital Sport Sciences). DIGISPORT est un projet d'École Universitaire de Recherche dans le domaine des sciences du sport, porté par l'Université Rennes 2, lauréat de la deuxième vague de l'appel à projets du programme d'investissements d'avenir PIA3.

Références

[1] Du Roy de Chaumaray, M., Marbac M. and Navarro, F. (2020). *Mixture of hidden Markov models for accelerometre data.* Annals of Applied Statistics, 14 (4), 1834-1855.

[2] Jacques, J., and Preda, C. (2014). *Functional data clustering : a survey.* Advances in Data Analysis and Classification, 8(3), 231-255.

[3] Marbac, M., Sedki, M., Biernacki, C. and Vandewalle, V. *Simultaneous semi-parametric estimation of clustering and regression.* arXiv 2012.14159.

[4] *NYC Department of Health and Mental Hygiene. Physical Activity and Transit Survey 2010-2011* https ://www1.nyc.gov/site/doh/data/data-sets/physical-activity-and-transit-survey-public-use-data.page