



Sujet de stage : mobilités et données manquantes

Origine du projet

Châteaubourg est la ville ayant la plus grosse densité au monde de données associées aux mobilités (voitures, vélos, piétons, poids-lourds). En effet, depuis trois ans, une quinzaine de capteurs [Telraam](#) mesure le flux de chaque mobilité toutes les heures, en distinguant les sens de parcours. Plus de 750 000 données ont été collectées et cela continue !

Le stage s'inscrit dans une collaboration entre l'association [Agis-Ta-Terre](#) qui installe et maintient ces capteurs et des enseignants-chercheurs de l'*Université de Rennes*.

Cette collaboration vise à développer des outils permettant de sensibiliser les citoyens sur l'usage des transports, mais aussi à aider les décideurs et les cabinets d'urbanisme quant à l'aménagement des territoires. L'objectif est de contribuer à un développement durable en mettant à disposition des outils libres et gratuits.

Présentation du projet

Cette base d'information fournit une vision claire des usages sur un temps long. Saisonnalité, impact du confinement, prix de l'essence... la ville, de 7 000 habitants, se révèle être un véritable *toy-model* pour la compréhension des mobilités à plus large échelle.

Depuis trois ans, un travail de post-traitement a été réalisé (sous *R*) afin de fournir des données sur le trafic dans la commune. Il s'agit d'une part d'une analyse statistique de ces données (*big-data*) et d'autre part de la création d'une interface publique et gratuite, [Mov-Around](#), permettant une lecture synthétique et efficace des analyses.

Comme les mesures sont réalisées toutes les heures depuis trois ans (pour certains capteurs) cet outil met à disposition des filtres (heures de la journée, type de véhicules, vacances) et des options de visualisation afin de faciliter l'interprétation.

Ce travail est actuellement valorisé via une interface *Rshiny* qui a été réalisée à cet effet et qui est déployée sur *GitHub*. Le [package d'interface](#) est propre aux données issues de Châteaubourg, il exploite les sources et modules déposés sur un [package de traitement statistique](#) qui vise à proposer des outils génériques de traitement des données Telraam. Comme souvent pour des données libres enregistrées en continu, les mesures ne sont pas toujours opérationnelles. Des paramètres extérieurs (luminosité, brouillard) ou humain (volets fermé, capteur déplacé) peuvent détériorer la qualité des données. Dans d'autres cas, le capteur peut être défectueux ou débranché par l'utilisateur. La série temporelle d'un capteur possède donc de nombreuses données manquantes. De plus, l'algorithme d'identification des mobilités utilisé par le capteur peut parfois être berné : il en résulte qu'une certaine proportion des données sont aberrantes.

L'objectif du stage vise à se concentrer sur ces données manquantes ou aberrantes afin de proposer des algorithmes de correction ou de filtrage permettant d'augmenter la fiabilité des

données exploitées. Ce travail reposera donc essentiellement sur le *package de traitement statistique*.

- Dans un premier temps il sera nécessaire de fournir une cartographie des données manquantes (en fonction des capteurs et des périodes)
- Il faudra alors proposer des algorithmes d'imputation de données manquantes, soit en *R*, soit en *Python*. L'attention sera portée à la fois sur l'efficacité des méthodes et leurs validation, mais également sur les temps de calcul sous-jacents.
- En ce qui concerne les données aberrantes, la tâche principal résidera dans l'identification de ces aberration et l'implémentation de ces test en fonction du type de mobilités.

Les travaux développés permettront d'améliorer la qualité des données, de mettre en oeuvre des alertes pour corriger les capteurs si besoin et éventuellement de contribuer au dépôt d'un package *R* sur le [CRAN](#).

Informations pratiques

Durée du stage : minimum 2 mois à partir de début mai

Le stage sera rémunéré. Vous serez en contact avec l'association *Agis-Ta-Terre* pour mieux appréhender les questions liées à cette problématique et la spécificité du territoire.

Contacts pour déposer votre candidature :

Véronique Thelen, veronique.thelen@univ-rennes.fr

Loïc Le Marrec, loic.lemarrec@univ-rennes.fr

Compétences

Bonne maîtrise de *R* et/ou *Python* et des outils de *datascience*. Une maîtrise de la programmation scientifique et de *GitHub* serait un plus.

Curiosité, capacité d'organisation, rigueur, autonomie et sens du travail en équipe. Ce dernier point est particulièrement important car plusieurs stagiaires travailleront en parallèle sur le projet.