

Prévision dans le réseau Vélib'

Étant un utilisateur régulier du service, je me suis aperçu d'un problème majeur. Assez souvent, certaines stations sont vides, d'autres pleines, on ne peut donc plus rendre le vélo emprunté. L'étude et la prédiction de l'évolution du réseau pourraient alors permettre aux utilisateurs de prévoir ces situations et les éviter.

Les services de vélos en libre-service (VLS), permettent un transport sans émission de pollution et participent à la désaturation des centres-villes. Ils semblent donc être une bonne solution pour le développement de transports alternatifs, un enjeu de taille pour le futur des villes.

Positionnement thématique (ÉTAPE 1) :

- *INFORMATIQUE* (Informatique pratique)
- *MATHEMATIQUES* (Géométrie)

Mots-clés (ÉTAPE 1) :

Mots-clés (en français) Mots-clés (en anglais)

<i>Vélos en libre-service</i>	<i>Bike-sharing system</i>
<i>Prévisions</i>	<i>Forecast</i>
<i>Algorithmes</i>	<i>Algorithms</i>
<i>Régression</i>	<i>Regression</i>
<i>Regroupement</i>	<i>Clustering</i>

Bibliographie commentée

En 2021, un article [1] est paru mettant en œuvre une technique de *clustering* des stations de vélos en libre-service (VLS). Ce regroupement des stations permet de considérer des propriétés communes au sein d'un groupe et de simplifier les échanges entre les groupes. On réduit alors largement le nombre de nœuds dans l'arbre représentatif. Dans cet article [1], le *clustering* se fait à l'aide d'une grille recouvrant l'ensemble de la surface d'effet du service de VLS, les stations sont alors dans des cellules de taille choisie et fixée. La prévision est ensuite faite statistiquement. L'application de cette méthode a été faite à Séoul.

Un peu plus tôt, en 2017, une autre étude [2] a été effectuée à San Francisco, dans laquelle la prévision est faite à l'aide des algorithmes de *Random Forest (RF)* et *Least-Squares Boosting (LSBoost)* et une variante le *Partial Least-Squares Regression (PLSR)* qui permet de limiter le nombre de modélisation à faire pour alléger le calcul. Les résultats obtenus étaient optimaux pour les prévisions à échéance de 15 minutes.

Toujours à San Francisco, en 2019, une autre recherche a été faite [3], dans laquelle a été utilisé

un modèle linéaire dynamique. Les chercheurs ont utilisé des modèles polynomiaux de premier et second ordre, intéressants pour leur simplicité. En effet, ces modèles reposent uniquement sur quelques équations linéaires, ce qui rend l'implémentation relativement facile. En outre, la fiabilité des résultats a été meilleure que celle des algorithmes *RF* et *LSB* en 2017 [2]. De plus, avec ce modèle, ils proposent une prévision à plus long terme, jusqu'à 120 minutes d'avance, qui est selon eux plutôt destinée aux services de rééquilibrage des stations. Cependant, ce type de prévision peut aussi servir à l'utilisateur pour envisager un moyen de transport alternatif.

En 2018, une équipe d'universitaires chinoise [4] a utilisé une technique dite de *Clustering hiérarchique*. Ils utilisent dans un premier temps un partitionnement spectral pour regrouper des stations, qui s'appuie sur un graphe pondéré des stations et "coupe" les arêtes pour créer des groupes. Cette méthode met aussi en jeu des matrices de similarités et leurs spectres. Ils appliquent ensuite l'algorithme de *gradient boosting*. L'idée est de multiplier les arbres de décision (qui sont individuellement faibles) et de garder la moyenne de leurs décisions (pour un résultat de qualité). L'application a été faite à New York, où ils ont regroupé les stations en 15 groupes. Et les résultats sont très satisfaisants même si les erreurs n'ont pas été quantifiées.

La même année, en Suisse, a été publiée une étude [5] dans laquelle est proposée une méthode de gestion de l'approvisionnement des vélos pour améliorer les conditions d'utilisation, limiter les stations pleines ou vides. Cette recherche a montré la supériorité du réseau de neurones vis-à-vis de la précision des résultats. Cependant, cette méthode est la plus difficile à mettre en place. Les meilleurs résultats ont été obtenus pour les prévisions à long terme. On peut faire alors un lien avec la méthode proposée par Mohammed H. Almannaa et al. [3]. Les deux modèles [5] et [3] sont donc *a priori* destinés aux services de rééquilibrage des stations mais peuvent aussi servir à l'utilisateur pour envisager un moyen de transport alternatif.

Une des difficultés principales est d'adapter les modèles proposés au système de données mises à disposition par la ville de Paris. Par exemple, les modèles comme celui de New York [4] utilisent le nombre d'arrivées et de départs de chaque station; d'autres modèles utilisent les données sous forme de vecteurs depuis la station de départ vers celle d'arrivée, comme à Séoul [1]. Il faut donc, quand c'est possible, en se basant sur les méthodes utilisées, réussir à les adapter.

Problématique retenue

Pour améliorer l'expérience utilisateur, il nous faut fiabiliser les prévisions d'évolution du réseau. Nous tâcherons donc de mettre en place un ou plusieurs modèles de prévision, avec différents paramètres (temporels, spatiaux, météorologiques), pour permettre une modélisation la plus fidèle à la réalité possible.

Objectifs du TIPE du candidat

L'objectif est donc dans un premier temps de comparer différentes façons de prédire l'évolution du réseau Vélib' en se basant sur les modèles présentés, en suivant une approche statistique puis algorithmique. Une fois des résultats satisfaisants obtenus, nous pourrions ajouter une

composante météorologique aux modèles de prévision. Nous pourrions aussi proposer une solution de gestion des vélos pour permettre un rééquilibrage, et éviter les saturations ou l'absence de vélos en station.

Références bibliographiques (ÉTAPE 1)

- [1] K. KIM : Spatial Contiguity-Constrained Hierarchical Clustering for Traffic Prediction in Bike Sharing Systems : *IEEE Transactions on Intelligent Transportation Systems*, vol. 23, no. 6, pp. 5754-5764, June 2022, doi: 10.1109/TITS.2021.3057596 <https://ieeexplore.ieee.org/document/9356471>
- [2] H. I. ASHQAR, M. ELHENAWY, M. H. ALMANNAA, A. GHANEM, H. A. RAKHA AND L. HOUSE : Modeling bike availability in a bike-sharing system using machine learning : *2017 5th IEEE International Conference on Models and Technologies for Intelligent Transportation Systems (MT-ITS)*, Naples, Italy, 2017, pp. 374-378, doi: 10.1109/MTITS.2017.8005700 <https://ieeexplore.ieee.org/document/8005700/>
- [3] MOHAMMED H. ALMANNAA, MOHAMMED ELHENAWY & HESHAM A. RAKHA : Dynamic linear models to predict bike availability in a bike sharing system : *International Journal of Sustainable Transportation*, DOI: 10.1080/15568318.2019.1611976 <https://doi.org/10.1080/15568318.2019.1611976>
- [4] S. FENG, H. CHEN, C. DU, J. LI AND N. JING : A Hierarchical Demand Prediction Method with Station Clustering for Bike Sharing System : *2018 IEEE Third International Conference on Data Science in Cyberspace (DSC)*, Guangzhou, China, 2018, pp. 829-836, doi: 10.1109/DSC.2018.00133 <https://ieeexplore.ieee.org/document/8411954>
- [5] S. RUFFIEUX, E. MUGELLINI AND O. ABOU KHALED : Bike Usage Forecasting for Optimal Rebalancing Operations in Bike-Sharing Systems : *2018 IEEE 30th International Conference on Tools with Artificial Intelligence (ICTAI)*, Volos, Greece, 2018, pp. 854-858, doi: 10.1109/ICTAI.2018.00133 <https://ieeexplore.ieee.org/document/8576131>

DOT

- [1] : mars - avril 2022 : découverte du site open.data.paris.fr, ces données donnent des idées de sujets, après des recherches bibliographiques je valide mon sujet de TIPE
- [2] : mai - juin 2022 : découverte de l'utilisation de l'API de Vélib'
- [3] : octobre 2022 : premiers résultats concluants avec le premier modèle (en utilisant des données de juillet) améliorations et utilisation de `np.load()` et du multiprocessing
- [4] : 23 octobre 2022 - 29 janvier 2023 : récupération des données principales
- [5] : janvier - février 2023 : implémentation du deuxième modèle
- [6] : mars 2023 : début de l'utilisation de l'algorithme de Random Forest, mais les résultats sont aberrants
- [7] : avril 2023 : aide reçue de Ricardo Ziegele un élève de Centrale Lyon pour l'utilisation de Random Forest, j'obtiens alors des résultats cohérents

[8] : mai 2023 : adaptation des algorithmes initialement prévus pour les places disponibles vers la prévision des vélos disponibles.