



HAL
open science

Prévision du nombre de passagers entrant dans un réseau de transport multimodal en cas d'évènements atypiques, à l'aide de méthodes d'apprentissage automatique

Florian Toque, Etienne Come, Latifa Oukhellou, Martin Trepanier

► To cite this version:

Florian Toque, Etienne Come, Latifa Oukhellou, Martin Trepanier. Prévision du nombre de passagers entrant dans un réseau de transport multimodal en cas d'évènements atypiques, à l'aide de méthodes d'apprentissage automatique. RFTM, 1ères Rencontres Francophones Transport Mobilité, Jun 2018, LYON, France. 3p. hal-01891340v1

HAL Id: hal-01891340

<https://hal.science/hal-01891340v1>

Submitted on 9 Oct 2018 (v1), last revised 9 Apr 2021 (v2)

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Proposition de communication pour les
1ères Rencontres Francophones Transport Mobilité (RFTM)
Lyon, 6-8 juin 2018

Titre :

Prévision du nombre de passagers entrant dans un réseau de transport multimodal en cas d'évènements atypiques, à l'aide de méthodes d'apprentissage automatique.

Auteur(s) :

*Florian Toqué, Université Paris Est - IFSTTAR, florian.toque@ifsttar.fr
Etienne Côme, Université Paris Est - IFSTTAR, etienne.come@ifsttar.fr
Latifa Oukhellou, Université Paris Est - IFSTTAR, latifa.oukhellou@ifsttar.fr
Martin Trépanier, Ecole Polytechnique de Montréal - CIRRELT, mtrepanier@polymtl.ca

Mots-clés :

Données massives · Apprentissage statistique · Carte à puce · Prévision de la demande · Transport Public · Prévision en cas d'évènements atypiques

Session(s) thématique(s) visée(s) :

Session 33 : Données massives en transport public : traitement et valorisation des données

Résumé :

Une large quantité de données est générée par les citoyens durant leurs déplacements (traces GPS, Wi-Fi, Bluetooth, données billettiques, réseaux sociaux, etc.). Ces données représentent une forte valeur descriptive de la mobilité urbaine, et peuvent permettre de développer des outils innovants en vue d'améliorer la qualité de service et de mieux organiser la mobilité des citoyens à l'échelle d'une ville. Les autorités publiques souhaitent orienter les pratiques de la mobilité urbaine de manière durable vers l'utilisation des transports publics et ainsi réduire la pollution provenant de l'utilisation des voitures individuelles. La compréhension de la mobilité urbaine au travers des réseaux de transport public est essentielle pour atteindre de tels objectifs. Nous avons focalisé nos travaux de recherche sur l'un des principaux capteurs de données de mobilité dans les réseaux de transport public, à savoir les données billettiques collectées à l'aide des cartes à puce. Dès 2005, Bagchi et White [1] ont étudié le rôle que ces données peuvent jouer dans l'analyse des comportements de mobilité et ont considéré leur potentiel pour compléter des sources de données existantes, telles que les données issues d'enquêtes. Depuis 2005, les données billettiques ont été utilisées dans de nombreuses études comme l'enrichissement et la complétion de données dans le but de valider l'estimation de destinations [2] ou pour inférer les motifs de déplacement [3]. Un autre domaine qui a été largement étudié porte sur l'analyse des comportements des usagers, au travers de

partitionnement automatique (ou clustering) des passagers dépendant de leurs activités dans un réseau de transport [4]. Plus récemment, on trouve des travaux dédiés à la prévision de la demande à la fois sur des horizons temporels courts et longs [5], [6]. La prévision de la demande dans les transports en commun est un réel challenge dans la mesure où la demande dépend d'une multitude de facteurs externes tels que le calendrier, la météo, des événements culturels, sportifs ou des ruptures de services. Ces différents facteurs sont difficiles à prendre en compte étant donné leur fortes irrégularités spatiales et temporelles. Dans cet article, nous nous attacherons au problème de la prévision court terme de la demande dans le cas d'événements atypiques (événements culturels ou sportifs, rupture de service) [7], [8]. Ces prévisions peuvent servir à optimiser l'allocation des ressources lors de ces événements et permettre aux usagers de réaliser une meilleure planification de leur trajet.

L'expérimentation a été menée sur un jeu de données billettiques réel d'une durée de 2 ans (2014-2015). Ces données ont été fournies par *Ile-de-France Mobilités* anciennement nommée la *Société des Transports d'Ile-de-France (STIF)*. Ce jeu de données représente le nombre de passagers entrant dans 30 stations du principal quartier d'affaire de Paris, *La Défense et ce, pour chaque quart d'heure*. Ces stations appartiennent à différents modes de transport, tels que le train (transilien), le métro, le RER (Réseau Express Régional) et le tramway. Les stations étudiées correspondent à 13 stations de la ligne de tramway T2, et 17 stations appartenant à 5 lignes de métro, RER ou train (transilien). Les stations ferrées hors tramway transportent plus de 215k passagers par jour tandis que la ligne de tramway transporte 35k passagers par jour. Dans le but de mettre l'accent sur la prévision de la demande durant des périodes ayant un comportement atypique, l'évaluation des modèles proposés est réalisé aussi bien, sur la totalité du jeu de données que sur un sous-ensemble de données qui contient uniquement des périodes avec des événements atypiques.

Nous proposons une approche issue de l'apprentissage profond, au travers d'un réseau de neurones récurrent. Cette approche a été comparée à la méthode des forêts aléatoires (Random Forest en anglais), largement utilisée en apprentissage statistique pour résoudre des problèmes de régression. Nous avons testé trois variantes du réseau de neurones récurrents (Gated Recurrent Unit, GRU) qui peuvent être décrites comme suit :

- Un modèle local qui considère indépendamment les différentes séries temporelles propres à chaque station
- Un modèle global, qui est un modèle multi-entrées/multi-sorties qui prédit simultanément toutes les séries temporelles des différentes stations
- Un modèle qui combine l'information obtenue de manière globale et locale à propos des différentes séries temporelles des stations

Les performances des modèles seront analysées à l'aide de différentes métriques d'évaluation telles que : la racine carrée de l'erreur quadratique moyenne (RMSE), le pourcentage de l'erreur moyenne absolue (MAPE), l'erreur moyenne absolue (MAE) et la fraction de variance non expliquée (FVU). Nous mènerons cette analyse individuellement sur chaque station puis de façon globale en considérant les performances sur l'ensemble des stations. Des résultats préliminaires sur la totalité du jeu de données montrent que les modèles de prévision locaux (Random Forest et GRU Local) obtiennent de meilleurs résultats que le modèle global (GRU Global) en général. En revanche, nous pouvons constater que le modèle GRU qui agrège l'information locale et globale obtient des résultats intéressants et prometteurs, notamment lors d'événements spéciaux. De tels résultats illustrent que ces types de modèles sont capables capturer les dépendances spatio-temporelles entre les différentes stations d'un réseau de transport pour des tâches de prévision de flux.

Remerciements :

Les auteurs souhaitent remercier l'autorité organisatrice de mobilité d'Ile-de-France (*Ile-de-France Mobilités*, anciennement *STIF*) pour avoir généreusement fourni les données billettiques.

Bibliographie

- [1] M. Bagchi and P. R. White, "The potential of public transport smart card data," *Transp. Policy*, vol. 12(5), pp. 464–474, 2005.
- [2] M. Munizaga, F. Devillaine, C. Navarrete, and D. Silva, "Validating travel behavior estimated from smartcard data," *Transp. Res. Part C Emerg. Technol.*, vol. 44, pp. 70–79, 2014.
- [3] S. G. Lee and M. Hickman, "Trip purpose inference using automated fare collection data," *Public Transp.*, vol. 6, no. 1, pp. 1–20, 2013.
- [4] M. K. El Mahrsi, E. Côme, L. Oukhellou, and M. Verleysen, "Clustering Smart Card Data for Urban Mobility Analysis," *IEEE Trans. Intell. Transp. Syst.*, vol. 18, no. 3, pp. 712–728, 2017.
- [5] M. Ni, Q. He, and J. Gao, "Forecasting the Subway Passenger Flow Under Event Occurrences With Social Media," *IEEE Trans. Intell. Transp. Syst.*, 2016.
- [6] N. Laptev, J. Yosinski, L. Erran, and L. Slawek, "Time-series Extreme Event Forecasting with Neural Networks at Uber," 2017.
- [7] F. C. Pereira, F. Rodrigues, and M. Ben-akiva, "Transport Arrivals Under Special Events Scenarios Using Data From the Web to Predict Public Transport Arrivals Under," vol. 2450, no. December 2017, 2015.
- [8] Y. Li, X. Wang, S. Sun, X. Ma, and G. Lu, "Forecasting short-term subway passenger flow under special events scenarios using multiscale radial basis function networks," *Transp. Res. Part C Emerg. Technol.*, vol. 77, pp. 306–328, 2017.