

**Proposition de communication pour les
2e Rencontres Francophones Transport Mobilité (RFTM)
Montréal, 11-13 juin 2019**

Titre :

Indicateurs de roulement de la clientèle des transports publics : une application pratique pour un algorithme de similarité des cartes à puce

Auteurs :

Vincent DIONNE, Conseiller en planification, Société de Transport de Laval, vdionne@stl.laval.qc.ca

Florian CYS, Conseiller en planification, Société de Transport de Laval, fcys@stl.laval.qc.ca

Anna DUSHINA, Technicienne en planification, Société de Transport de Laval, adushina@stl.laval.qc.ca

Mots clés :

Données massives, cartes à puce, fusion de données

Résumé long :

1 – Introduction

Les indicateurs de roulement de la clientèle sont essentiels pour évaluer la performance d'une entreprise, et les sociétés de transport n'y font pas exception. Typiquement, elles se fient aux attitudes et comportements déclarés dans les sondages de satisfaction de la clientèle (Perk et al., 2008; Trépanier et Morency, 2010) faute de données nécessaires pour suivre en continu le comportement observé (Trépanier et al., 2012). À cet égard, les données issues de validations des cartes à puce (CAP) présentent un potentiel unique pour mieux comprendre le roulement de la clientèle du transport en commun en termes de rétention, acquisition, attrition et variation (Blake et al., 2017).

Toutefois, une limitation importante subsiste. Les usagers des transports en commun de la région de Montréal ont l'obligation de remplacer leur CAP après une période de 1, 2, 4 ou 7 ans (dépendamment de la catégorie d'usager). Ainsi, même s'il est possible de suivre le comportement enregistré avec une CAP, le suivi d'un client au-delà du remplacement de cette carte est difficile, voire même impossible (Chu, 2015). Afin de pallier à cette limitation, la Société de Transport de Laval (STL) a développé et mis en œuvre un algorithme qui permet d'associer certaines CAP remplacées avec des cartes nouvellement apparues sur le réseau, sur la base de similarité de leurs comportements spatio-temporels. Ce travail est notamment inspiré des recherches sur l'enrichissement des données CAP afin de dériver les lieux d'activités des détenteurs de carte et leur comportements spatio-temporels (Chu et Chapleau, 2014), ainsi que des modèles développés pour mesurer la similarité des comportements de passagers à l'aide des données CAP (Zheng, 2014; Faroqui, 2018; Gaemi, 2015).

Notre présentation porte sur l'application d'un algorithme de similarité de CAP et de son impact sur les indicateurs de roulement de la clientèle de la STL. En premier lieu, nous aborderons le développement de l'algorithme qui a été appliqué sur 3 ans de données CAP. Ensuite, nous illustrerons l'impact de cet algorithme sur les mesures classiques de roulement de la clientèle.

2-Concepts et base théorique

2.a. Roulement de la clientèle

L'augmentation du nombre d'usagers du transport en commun figure parmi les enjeux fondamentaux auxquels font face les grandes aires urbaines, et les stratégies de marketing prennent de plus en plus de place dans les

leviers des autorités de transport pour y répondre. Une approche qui permet d’avoir une vue d’ensemble du roulement de la clientèle s’inspire de la théorie « *leaky bucket* » et postule que pour mieux comprendre l’évolution de la clientèle, il est essentiel d’analyser le taux d’acquisition de nouveaux clients en même temps que les taux d’attrition et de rétention des clients existants (Blake et al., 2017).

En supposant la période d’un an comme unité pour mesurer ces taux, il faut évaluer le nombre de clients actifs au début de la période (C_D), le nombre de clients actifs à la fin de la période (C_F), le nombre de clients perdus au courant de la période (C_P), et le nombre de nouveaux clients acquis au courant de la période (C_N). Ainsi,

$$C_F = C_D - C_P + C_N$$

$$\text{Taux de rétention} = (C_F - C_N) / C_D$$

$$\text{Taux d'attrition} = 1 - \text{Taux de rétention} = C_P / C_D$$

$$\text{Taux d'acquisition} = C_N / C_D$$

Cette approche est loin d’être universellement utilisée dans le domaine des transports en commun (Blake et al., 2017), et les rares études connues se basent sur les données des sondages (Mason et al., 2011). Quant aux études basées sur les données issues de validation des cartes OPUS, elles adoptent une approche différente et portent surtout sur les habitudes de comportement individuelles ou par segment de marché (Chu, 2015) ou bien sur la relation entre la fidélité des clients et les facteurs individuels ou les caractéristiques de l’environnement bâti (Trépanier et Montmorency, 2010; Trépanier et al., 2012). La STL propose donc une application additionnelle des données CAP pour la mesure de la performance de son service.

2.b. Algorithme de similarité des cartes à puce

Les données de validations des CAP à bord des autobus comportent des identifiants uniques attribués à chaque carte et ainsi permettent d’établir un lien entre les embarquements. Les identifiants sont toutefois anonymes, et donc n’offrent pas de référence directe vers l’identification personnelle de l’usager qui détient une carte. La perte de la CAP ou bien son arrivée à échéance conduisent alors l’identifiant à disparaître du jeu de données. Ainsi même si la disparition d’une CAP ne signifie pas nécessairement l’arrêt de l’usage du service par le client, du point de vue de l’agence de transport, il n’est plus possible d’associer les validations effectuées par ce même client à l’historique de son utilisation avant le changement de carte. La seule façon de rétablir ce lien se base sur l’analyse des habitudes spatio-temporelles d’utilisation des CAP.

Les méthodes d’enrichissement de données permettent de faire correspondre un arrêt à une validation à bord du bus et ainsi obtenir une référence spatiale et temporelle de l’embarquement. Nous définissons alors sur cette unité spatio-temporelle le concept de prisme spatio-temporel, PST, adapté de Faroqui et al. Un PST, construit itérativement, regroupe l’ensemble des validations telles qu’elles soient proches dans le temps et l’espace. Au final, chaque PST est qualifié par son barycentre et son poids, égal à la somme des validations qu’il intègre (voir Figure 1).

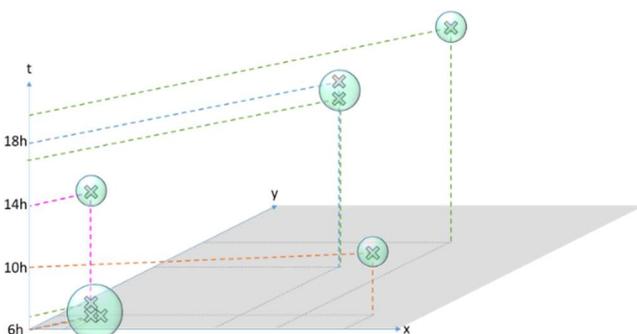


Figure 1. Représentation des prismes spatio-temporels (PST)

Chaque CAP possède alors un ensemble de PST. La similarité de deux supports peut alors se mesurer à l'aide de la somme de la distance pondérée par le poids pour chaque PST de la CAP 1 vers la CAP 2 et pour chaque PST de la CAP 2 vers la CAP 1, divisée par la somme totale des poids des PST.

Et où la distance entre 2 PST s'établit par la distance euclidienne spatio-temporelle entre les 2 lieux plus une pénalité basée sur les réseaux, les poids et la variation du temps par rapport à l'espace.

Une CAP apparue est similaire à une CAP disparue si :

- La disparition de la CAP précède de moins de 135 jours l'apparition de la suivante.
- Les titres en usage sur la CAP née sont compatibles avec les titres en usage sur la CAP disparue.
- Le score de similarité est inférieur à un certain seuil (ScoreMax)

Pour chaque CAP apparue, aucune à plusieurs cartes similaires disparues sont identifiées. Individuellement, le plus petit score de similarité doit être choisie, mais sur l'ensemble des cartes, il importe d'assurer que l'association soit unique. Par itération, les paires de CAP disparue et apparue sont sélectionnées en ordre des minima d'écart jusqu'à constituer un ensemble de toutes les similarités de supports possibles.

3- Données et résultats

L'algorithme a été appliqué sur 3 ans (2015-2016-2017) de transactions CAP sur le réseau de la STL. Celles-ci représentent environ 80% des transactions de paiement totales à l'embarquement. Sur 3 ans, 194 432 nouvelles cartes ont été identifiées. De cet ensemble de départ, seules les cartes avec au moins 100 validations ont été retenues afin d'assurer un portrait suffisant des habitudes de ces cartes. Sur trois ans, cela représente donc 136 528 cartes. Pour la même période, 141 049 cartes de plus de 100 validations ont cessé d'être utilisées. Les PST de chaque carte apparue et disparue ont été déterminés à partir des 100 premières ou des 100 dernières validations (total de 28 millions de transactions).

Ne disposant d'aucun moyen d'estimer l'erreur des résultats de l'algorithme à l'aide de données réelles de changement de carte, un outil de validation manuelle a été développé et sera présenté. La validation consiste à évaluer la vraisemblance de l'association basée sur la distribution spatio-temporelle des deux cartes associées. 134 associations ont été manuellement validées. Cet outil a permis de déterminer un score maximal pour qu'au moins 90% des associations soient valides selon l'analyse manuelle.

Au final, sur les 136 528 cartes nées, 45 035, soit un tiers, ont été associées.

4 – Impact de l'algorithme sur les mesures de roulement de la clientèle

Dans le cadre de l'évaluation de la performance de l'entreprise, nous avons évalué l'évolution du nombre de clients actifs de la STL et calculé les taux de rétention, d'attrition et d'acquisition de ses clients, et ce, avec les données brutes de validation CAP ainsi qu'avec les données issues de l'algorithme d'association des CAP.

Une comparaison des résultats de ces calculs nous a permis de constater que, tandis que les tendances (à la hausse ou à la baisse) affichées étaient assez semblables pour les deux types de données, il y a une différence marquée d'échelle : pour le calcul mensuel glissant du taux de rétention, les résultats oscillaient autour de 29% pour les données brutes, tandis que pour les données des CAP associées à l'aide de l'algorithme de similarité, le taux se situait plutôt autour de 40% (voir Figure 2).

Un autre effet marqué (et attendu) de l'algorithme est le lissage de l'effet saisonnier, et en particulier, la réduction des sommets observés aux mois de septembre et octobre et attribuables à la date limite de remplacement des CAP étudiantes au 31 octobre de chaque année.

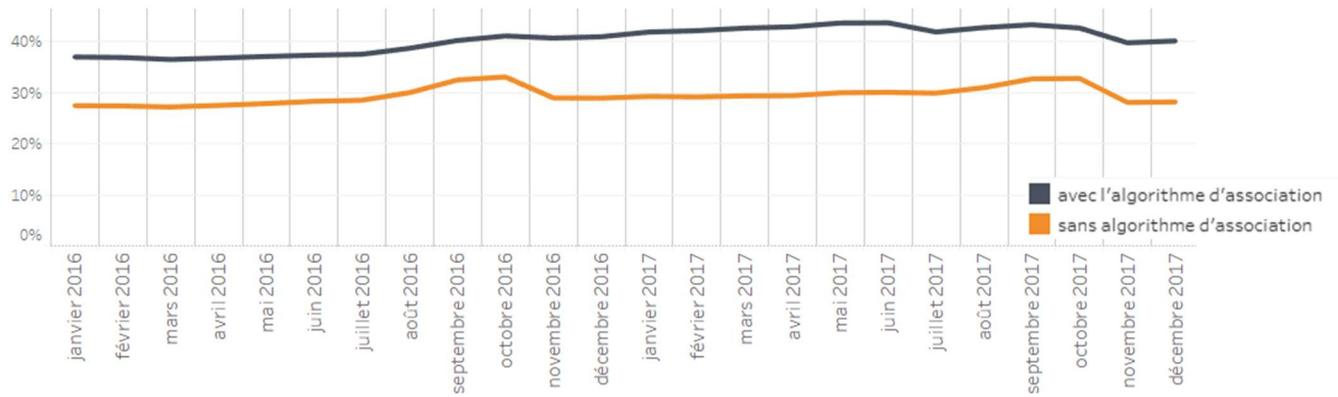


Figure 2. Taux de rétention des clients de la STL sur une période d'un an (calcul glissant)

Pour ce qui est du taux d'acquisition des nouveaux clients, nous y voyons également un écart univoque et constant entre les résultats fondés sur les données brutes et ceux obtenus à l'aide de l'algorithme de similarité : la proportion des clients que l'on considère comme nouveaux est, en moyenne, 11% moins élevée si l'on utilise les résultats de l'algorithme (voir Figure 3).

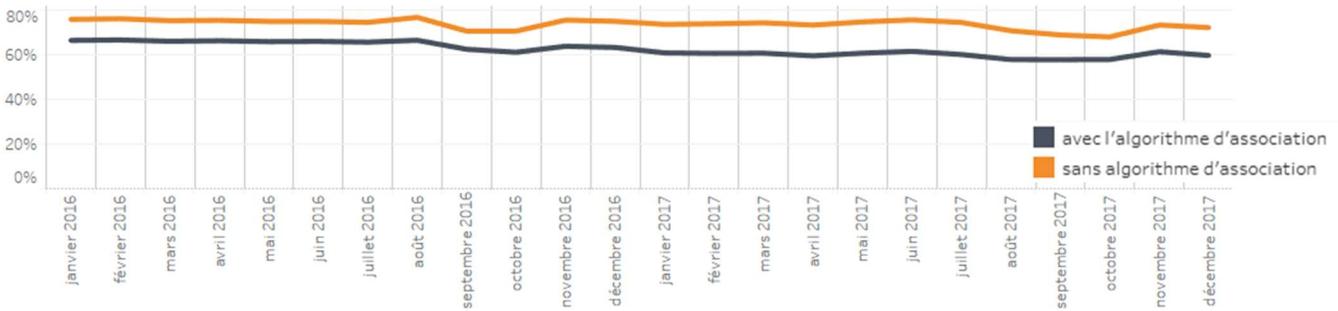


Figure 3. Taux d'acquisition des clients de la STL sur une période d'un an (calcul glissant)

5 – Conclusions

L'utilisation des données CAP est une avenue prometteuse pour évaluer le roulement de la clientèle du transport en commun.

L'effet de remplacement des CAP est (au moins partiellement) mitigé par le développement et la mise en œuvre d'un algorithme d'association des CAP basé sur la similarité des comportements spatio-temporels. L'impact de cet algorithme sur les taux de rétention et d'acquisition est constant et non-négligeable.

L'algorithme permet de faire une association pour 33% des cartes apparues sur une période de 3 ans. L'absence de données réelles pour valider les associations demeure un enjeu, mais l'utilisation d'un outil manuel de validation nous permet d'assurer la cohérence des résultats.

La démarche ouvre la porte pour l'analyse du comportement des clients sur une plus longue période, ainsi que pour l'étude du lien entre sur la fidélité de la clientèle et les facteurs comme la qualité du service vécu et les nouveaux produits tarifaires. Un éclatement des résultats par type de client (Ex. : étudiant, personne âgée) pourra alimenter des stratégies de marketing plus ciblées au niveau du changement de comportement des clients.

6 – Références

- Blake, P., G. Currie, A. Delbosc, C. Lowe. "Customer Churn: The Missing Link in Public Transport Marketing." *Australasian Transport Research Forum Proceedings*. 2017
- Chu, K. Two-year worth of smart card transaction data—extracting longitudinal observations for the understanding of travel behaviour, *Transportation Research Procedia*, 2015
- Chu, K. et R. Chapleau. Augmenting Transit Trip Characterization and Travel Behavior Comprehension, *Transportation Research Record*, 2014
- Faroqi,H., Mesbah,M., Kim,J., et Tavassoli A.,A model for measuring activity similarity between public transit passengers using smart card data,*Travel Behaviour and Society*,Volume 13,2018
- Mason, Adam, John Segal, and Ben Condry. "Market churn in the British rail passenger commuter and leisure markets." *European Transport Conference 2011, Association for European Transport, Transportation Research Board*. 2011.
- Perk, V., J. Flynn, et J. M. Volinski. *Transit ridership, reliability and retention*. No. NCTR-776-07. 2008.
- Trépanier, M., K. Habib, C. Morency. "Are transit users loyal? Revelations from a hazard model based on smart card data", *Canadian Journal of Civil Engineering*, 2012
- Trépanier, M., et C. Morency. "Assessing transit loyalty with smart card data." *12th World Conference on Transport Research, Lisbon, Portugal*. 2010.
- Zheng, B., K. Zheng, M. A. Sharaf, X. Zhou et S. Sadiq, "Efficient Retrieval of Top-K Most Similar Users from Travel Smart Card Data," *2014 IEEE 15th International Conference on Mobile Data Management*, Brisbane, QLD, 2014, pp. 259-268
- Ghaemi, M., B. Agard, P. Nia, M. Trépanier. Challenges in Spatial-Temporal Data Analysis Targeting Public Transport, *IFAC-PapersOnLine*, Volume 48, Issue 3, 2015