A density-based smart card data classification algorithm applied to land use analysis and infrastructure performance measure

Li He[1], Martin Trépanier[1], Bruno Agard[1], Marcela Munizaga[2], Benjamin Bustos[2]

1. Polytechnique Montréal; Interuniversity research centre on enterprise networks logistics and transportation. 2. Universidad de Chile.

Background

Data from smart card fare collection systems is very useful to public transit planners [Pelletier and al., 2011]. This data can be used for better understanding passenger travel behavior and public transit demand, so that public transit authority can offer a better service to passengers. In this presentation, we present a method to use smart card data to analyse land use and measure the performance of transport infrastructure.

Literature review

Several articles analyze passenger travel behavior for each transaction [Trépanier and al., 2007]. In the field of data mining, a series of researches show the application of data mining in characterization of public transit smart card users' behaviors: K-means has been used to classify public transit users' behaviours into a few clusters [Agard et al., 2006]. Some works have been down to deal with spatio-temporal behaviours of public transit users [Ghaemi and al., 2015; Ghaemi and al., 2016]. Then users' daily spatio-temporal information has been treated as time series, and three classification algorithms have been developed to classify spatio-temporal daily behaviours of public transit users [He et al., 2018; He et al., 2019].

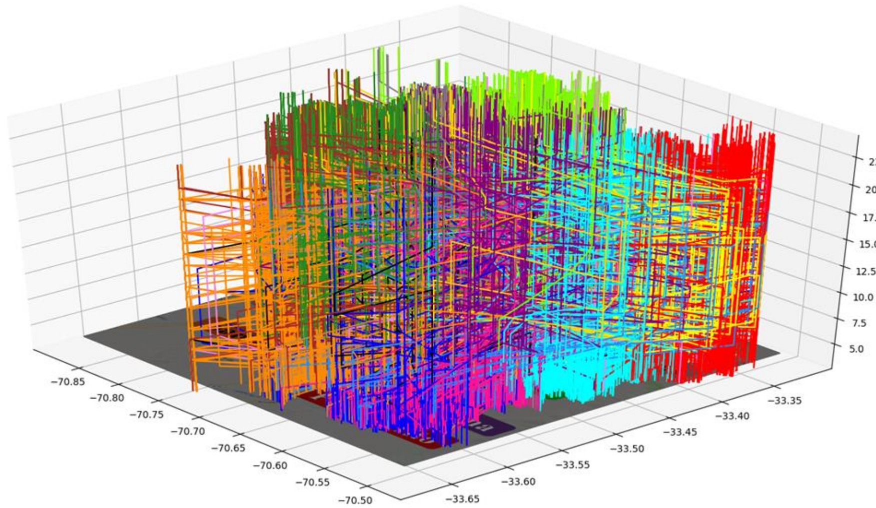

Figure 1: Result of spatio-temporal public transit users' behaviours. X and Y axis: Santiago map. Z axis: Hours for one day [He et al., 2019]

Problematic and objectives

Many methods have been developed to classify users' behaviours. However, few classification algorithms have been developed to classify zones in a city. Then, even though the DBSCAN (Density-based spatial clustering of applications with noise) has been applied to identify public transit regular passengers, the density-based classification method has few been used in this issue (zone classification). Therefore, the general objective of this article is to develop an algorithm based on density to classify zones of public transit users.

With this new algorithm, some applications would be done. First, the land use of a city could be "derived" by using this method. Then, some analyses would be done based on the method developed to measure infrastructure implementation performance. The change of public transit users' behaviours would be measured before and after the implementation of subway line 6 of Santiago and Rapibus (BRT) of Gatineau.

Methodology

Regarding the method of land use analysis, the result of the spatio-temporal classification can be used to divide territory into land use zones (residential or commercial/study) based on the public transport transaction location and time. As mentioned in Figure 1, the spatio-temporal path can (for example) be cut at 3 a.m. in the morning to locate the home of each user. Next, we calculate the user density of each cluster using the kernel estimate, and then choose the cluster with the maximum density to represent a zone.

Regarding the method of infrastructure implementation performance measure, for example we calculate the density of users' first transaction time period (example: 8:56 is in the period of 8:30-9:00), then choose the cluster with the maximum density to represent a zone.

Implementation

The algorithm is programmed in python.

(1) For land use analysis, the data of 31st July 2017 of TranSantiago is used to test the algorithm.

(2) To measure the performance of line 6 of Santiago, the data of 9th April 2018 of TranSantiago is used.

(3) To measure the performance of Rapibus of Gatineau, the data in a week of September of the year 2013 and 2014 of STO (Société de Transport de l'Outaouais) is used. The Rapibus (BRT) system is implemented in December 2013.

Result and analysis

Land use analysis

Figure 2 is the result of land use analysis. The density of user groups (of spatio-temporal behavior) divides the city into 16 zones. It is also interesting to see that:

(1) The metro terminals are often close to the limit of two zones. For example, the north terminal of line 2 (yellow line) is near the border of green and purple lawn zones. This means that the behavior of smart card users changes a lot after leaving the area covered by subway.

(2) Subway lines may extend a zone. For example, the pink group (southeast) includes two very narrow areas with subway stations. User behaviors become more similar because of the subway.

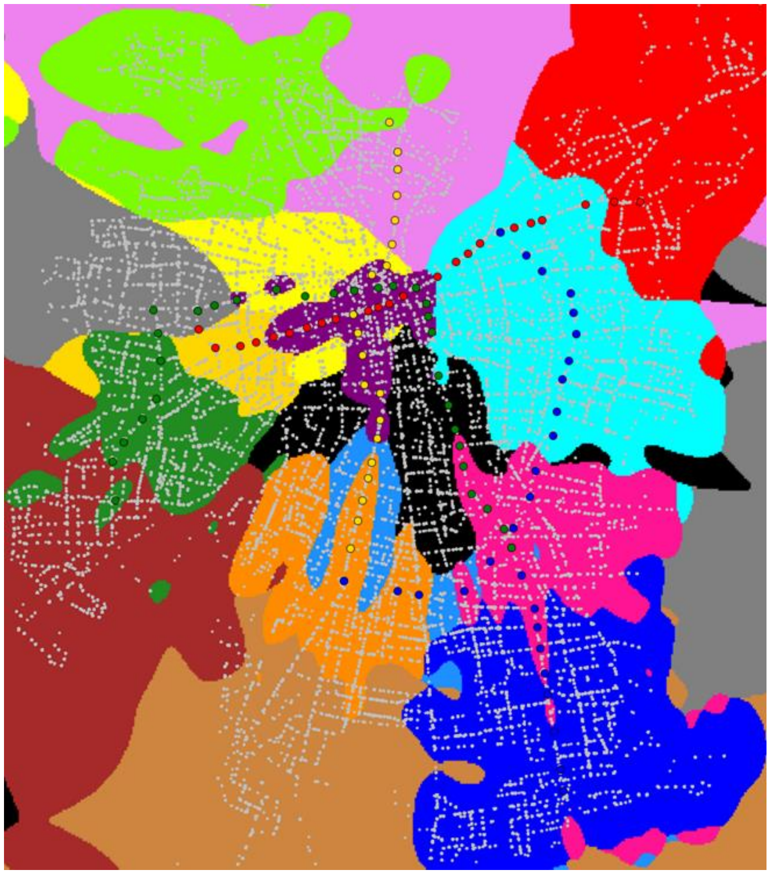

Figure 2: Land use analysis

Performance of line 6 of Santiago

The result of zone classification based on the time of first transaction is presented in the Figure 3. The bandwidth is parameter that represents that if a one-stop transaction has an impact on a stop that is very far away. When choosing a wide bandwidth, we can see more aggregated user behaviors (Figure 3), while choosing a low bandwidth, user behaviors are more disaggregated (Figure 4).

In the Figure 4, the dominating first transaction time in the zone of last west 3 stations of line 6, is much later than that of the zones far from these stations, we may infer that the residents near these stations leave home later after the implementation of line 6.

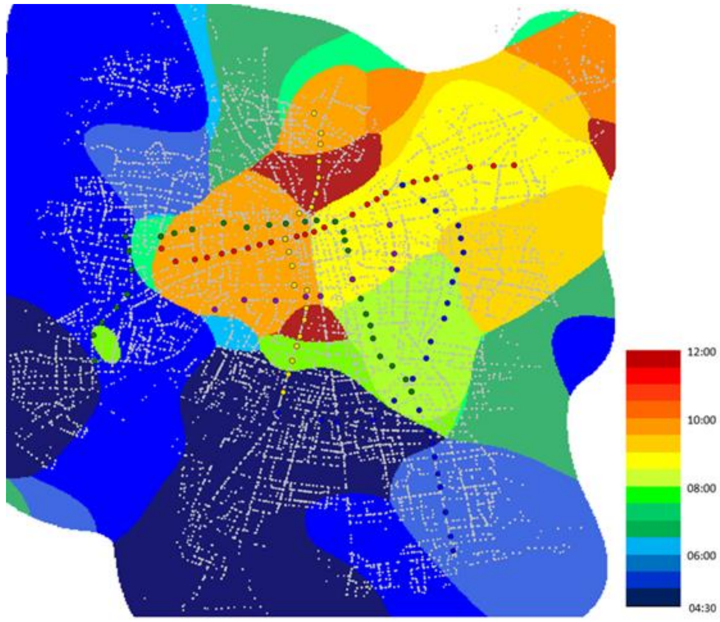

Figure 3: Classification of zones based on the first transaction time – aggregated

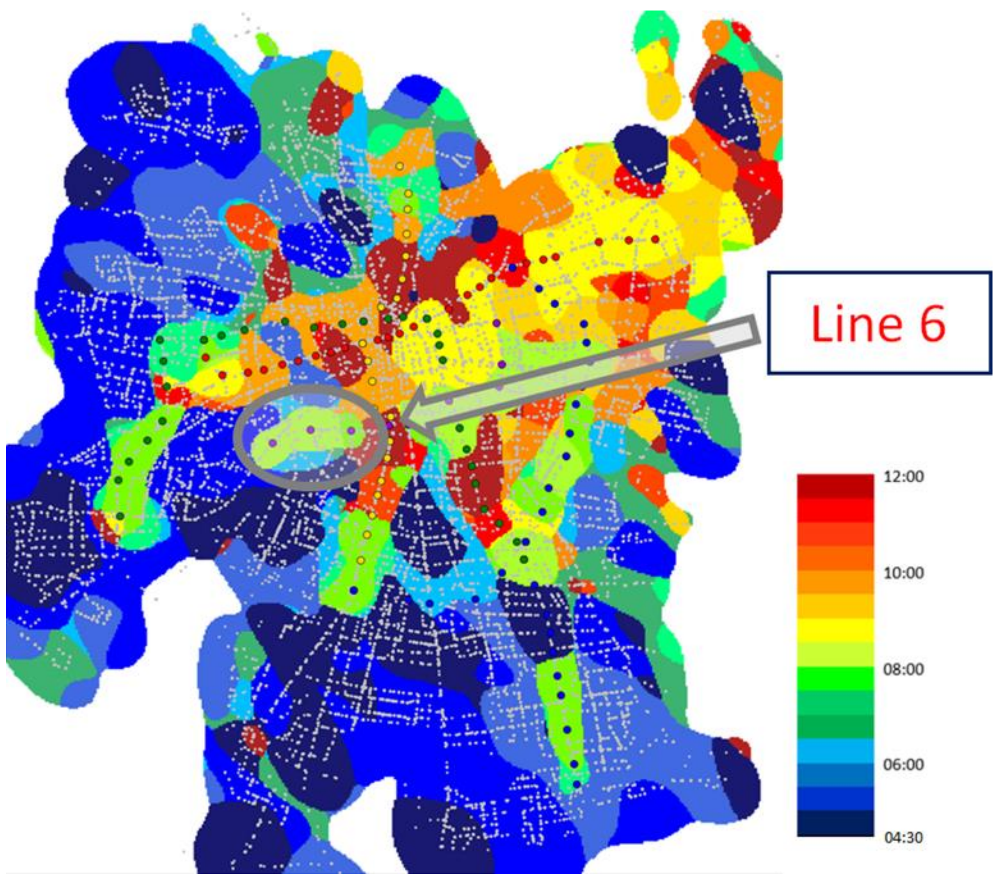

Figure 4: Classification of zones based on the first transaction time – disaggregated

Performance of Rapibus

Based on Figure 3, similarly, a zone classification based on the STO time classification groups is made. In this way, we can make a comparison before and after the implementation of BRT (Rapibus).
Different bandwidths have been also tried (0.005 and 0.01). For both cases, passengers from certain stops travel later. However, others travel earlier. The time of first transaction of a day does not change in a significant way after the implementation of the Rapibus.

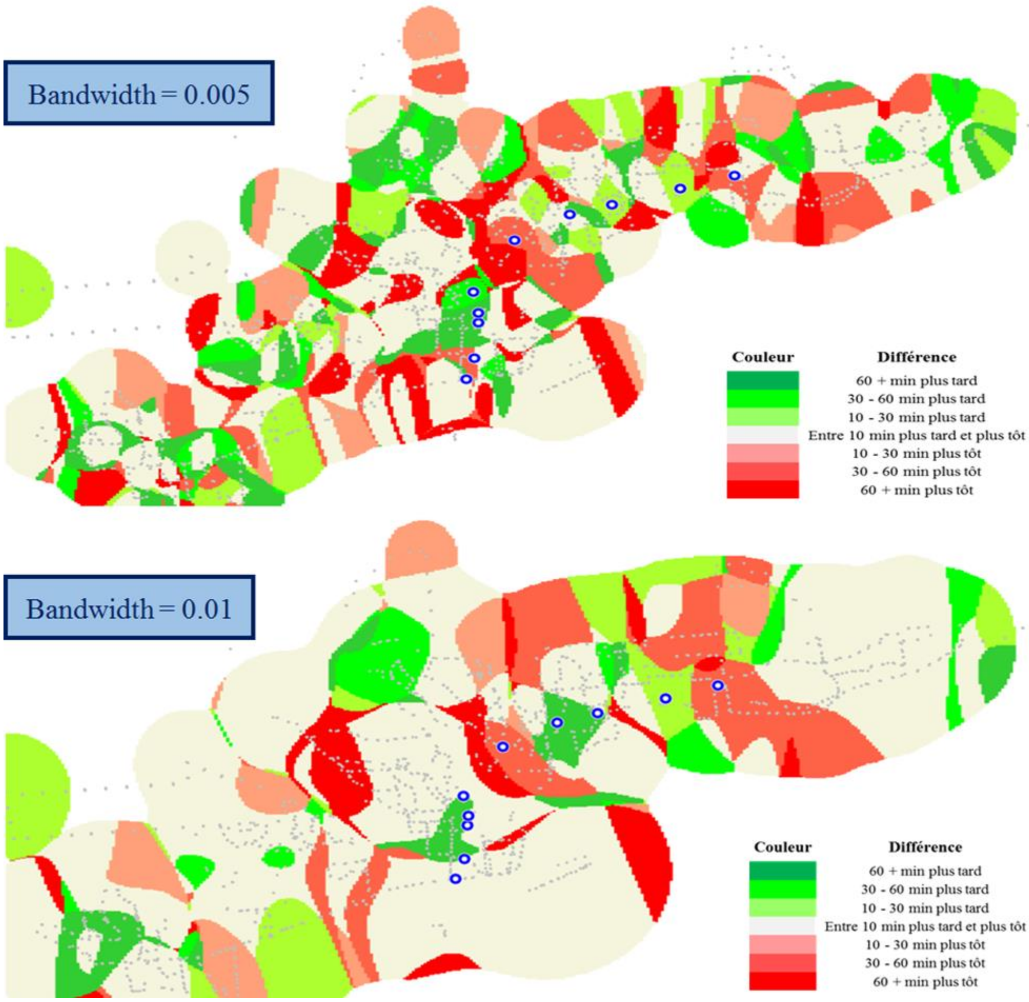

Figure 5: Time difference of first transaction before and after Rapibus implementation

Key words

Public transit smart card data, data mining, density-based algorithm, land use analysis, infrastructure performance measure

Reference

Agard, B., Morency, C., & Trépanier, M. (2006). Mining public transport user behavior from smart card data. *IFAC Proceedings*. Volumes, 39(3), 399-404.

Ghaemi, M. S., Agard, B., Nia, V. P., & Trépanier, M. (2015). Challenges in Spatial-Temporal Data Analysis Targeting Public TransportÕ. *IFAC-PapersOnLine*, 48(3), 442-447.

Ghaemi, M. S., Agard, B., Trépanier, M., & Partovi Nia, V. (2016). A Visual Segmentation Method for Temporal Smart Card Data. *Transportmetrica A: Transport Science*, 13(5), 381-404.

He, L., Agard, B., & Trépanier, M. (2018). A classification of public transit users with smart card data based on time series distance metrics and a hierarchical clustering method. *Transportmetrica A: Transport Science*, 1-20.

He, L., Trépanier, M., Agard B., Munizaga M., Bustos B. (2019). Spatio-temporal classification of transit user behavior with smart card data. Presented at *Workshop of International Network on the Use of Passive Data from Public Transport*, Washington, DC.

Pelletier, Marie-Pier, Martin Trépanier, and Catherine Morency. "Smart card data use in public transit: A literature review." *Transportation Research Part C: Emerging Technologies* 19.4 (2011): 557-568.

Trépanier M, Tranchant N, Chapleau R (2007). Individual trip destination estimation in a transit smart card automated fare collection system. *Journal of Intelligent Transportation Systems* 11(1):1–14.