ASSESSING PUBLIC TRANSPORT TRAVEL BEHAVIOUR FROM SMART CARD DATA WITH ADVANCED DATA MINING TECHNIQUES

(13 pages)

Bruno AGARD
Polytechnique Montréal, dept. Mathematics and Industrial Engineering
2900, boul. Édouard-Montpetit, Montréal, QC, Canada, H3T 1J4
bruno.agard@polymtl.ca

Vahid PARTOVI NIA
Polytechnique Montréal, dept. Mathematics and Industrial Engineering
2900, boul. Édouard-Montpetit, Montréal, QC, Canada, H3T 1J4
vahid.partovi-nia@polymtl.ca

Martin TRÉPANIER
Polytechnique Montréal, dept. Mathematics and Industrial Engineering
2900, boul. Édouard-Montpetit, Montréal, QC, Canada, H3T 1J4
mtrepanier@polymtl.ca

ABSTRACT

Smart card automated fare collection system are continuously gathering data on transactions made in public transport systems. While this data is not mainly aimed to planning purposes, it can provide very useful evidences on travel behaviour variability over space and time. The aim of this paper is to present innovative data mining techniques for grouping and characterising public transport users. This study uses data from the Gatineau, Québec, public transport agency smart card data. Covering a one-year periods, 9.4 millions transactions were recorded on the 200 bus network. An innovative distance calculation technique is proposed to apply the k-means clustering method. This distance measurement better captures the likelihood of travel behaviour between two different cardholders. The method has been applied to a set of 4.45 millions cards-day to uncover 3 clusters of travel behaviour among the users.

Keywords: public transport, smart card, data mining, travel behaviour

INTRODUCTION

The vast majority of public transport systems around the world are schedule-based. Schedules are convenient both for the public transport user and for the public transport authority, because it helps to operate the service while maintaining a reliability feeling for the customers. Most of the time, service providers operate on the same schedule for all the weekdays from Monday to Friday, and maintain distinct schedules for Saturdays and Sundays, plus Holidays. This assumes that the public transport user follows the same travel behaviour during weekdays. It could be true for five-days, regular workers who commute back and forth from work. However, society is constantly changing and more people now work only 4 days while other telework once or twice a week. In addition, there are an increasing number of senior citizens with non-regular schedule. There is maybe a need, in the near future, to adapt public transport schedule to the irregularity of the transport behaviours. However, this irregularity is still to be proven, and many problems, which will not be discussed here, would have to be overcome before having separate schedules for each day of the week.

This paper proposes the use of advanced data mining techniques to assess the regularity of public transport behaviour from time stamped smart card transactions. By "advanced techniques", we mean using a blend of judicious database processing and in-depth clustering method using a special distance function to obtain clusters of user behaviours that can be put back in the context of daily mobility. By analysing these clusters, one could understand the different categories of users, especially those who have a regular pattern of travel, compared to those who have not. The objectives of the paper are twofold: first, improve the processing of smart card data and the usage of data mining techniques in transport; second, to help public transport planners to assess their customer behaviours, hoping it would bring them to provide a service more suitable to the demand.

The structure of the paper is straightforward. First, a literature review recalls some work done in the field of smart card data analysis for public transport and in the use of data mining techniques. Then, the methodology is presented: the case study and the information system are presented, and the mathematical methods are described. The presentation of the results follows, and the paper is concluded with a discussion on the contribution and the limitations of the proposed approach.

BACKGROUND

Since the advent of smart card automated fare payment systems in public transport, many efforts are made to use the transaction data for public transport planning purposes. This is a continuous challenge because the payment systems are not intended for this task at first. Their design is made to collect revenue by registering every transaction made in the stations and the vehicles of the public transport network.

Smart card in public transport planning

The use of smart card data in public transport is thoroughly described in the review paper by Pelletier et al. (2011). They identified three categories of study, depending on the organizational level focussed in the process, may it be strategic, tactical or operational.
At the strategic level, Park and Kim (2008) have looked at historical data to create a future demand matrix for the Seoul network. Chu and Chapleau (2008) also addressed network design issues by building a spatio-temporal portrait of the Gatineau network from smart card transactions. Bagchi and White (2005) performed turnover analyses to estimate the number of card users that entered or quitted the system during a certain period of time. The concept has been developed furthermore by Trépanier et al. (2012) who proposed a hazard model to assess the loyalty of smart card users.
Tactical level studies are mostly related to service adjustment and public transport user journey analysis. Usually, in the smart card system, there is no information about the alighting point of the users along a route, so this has to be estimated. Trépanier et al. (2007) proposed an algorithm based on the sequence of trips within a journey. The method has been improved by Munizaga et al. (2010) in the case of subway trips. The construction and examination of travel patterns has been made by Bacghi and White (2005) and Seaborn et al. (2009). The development of origin-destination matrices form smart card data has also been proposed by several research teams (Gordon et al. 2013, Munizaga et al. 2010, Zhao et al. 2007). Devillaine et al. (2012) studied the activities of public transport users with the help of smart card data from Santiago, Chile and Gatineau, Quebec.
Operation-level studies are dedicated to day-to-day service issues such as daily key performance indicators calculation, data error correction and user information. Reddy et al. (2009) calculated several operational statistics available at individual level with the help of magnetic card data, similar to smart card data. Trépanier et al. (2009) have calculated multiple supply and demand indicators like passenger-kilometres, vehicle-kilometres, average speed and adherence measures from smart card data for the Gatineau, Quebec, network. Chapleau and Chu (2007) developed a method based on travel patterns to correct errors in smart card data.

Clustering techniques

Clustering techniques offer the opportunity to segment data in different groups where data in the same group share some similarities while data in different groups are less similar. Many segmentation methods are available; Xu and Wunsch II (2005) classified them in different general categories: distance and similarity measures, hierarchical, squared error-based, mixture densities-based, graph theory-based, combinatorial search techniques-based, fuzzy-based and others. Each category is then divided according to its philosophy for segmenting the data. Most of the methods have difficulties managing really large sets of data due to their computational complexity and few of them are used on practical problems because of not user-friendly parameters to fix.
In the area of public transport many analyses with segmentation relies on k-means which permit to compute relatively large sets of data and that requires few parameters to be fixed, besides one important parameter is the number of segment to produce. Recent studies

propose mathematical optimal number of clusters based on statistic gap (Tibshirani et al. 2001), once again on practical analysis the "optimal" number of clusters may lead to non-manageable results for decision making and user expertise is often require to discriminate between different possible number of clusters (Morency et al., 2007, Lathia et al., 2012).

Smart card and data mining

Data mining is a collection of techniques and tools dedicated to the discovery of non-trivial, implicit, previously unknown, and potentially useful and understandable patterns from large datasets (Anand and Büchner 1998). These methods are well suited to smart card automated fare collection system data, because they collect huge quantities of data (millions of transactions daily in the case of large cities). In addition, the travel behaviour of users can be modelled in different patterns according to several dimensions like the day of the week, the time of departure, the route choice and the fare categories. This makes it difficult to analyse by traditional statistical methods, especially if one may look at the behavioural change of users from time to time.
Morency et al. (2006) used data mining clustering techniques (essentially the K-means algorithm) to identify and analyse groups of users according to the behavioural patterns of their daily travel. They discussed about the variability of public transport user patterns in a subsequent work (Morency et al. 2007). The k-means technique has also been used by Ma et al. (2013a,b) for both the categorization of travel patterns and the extraction of the origin information about trips.

METHODOLOGY

This section describes the data and the clustering technique used in this study.

Information system

Data from this study comes from the Automated Fare Collection Smart Card System from the *Société de transport de l'Outaouais* (STO), based in Gatineau, Quebec. The STO authority operates a contactless smart card system since 2001 in its 200-buses network. Each day, the system collects data on every transaction made when public transport users board the buses. The dataset covers a one-year period between January 1st and December 31st, 2008. A total of 9.4 millions transactions were made in the network by 52,825 smart cards. For each transaction, the following attributes are available:

- Date and time of the boarding transaction;
- Card number and fare type;
- Route number and direction;
- Vehicle and driver numbers;
- Stop number at boarding.

For this specific study, only card number, fare type, date and time of transaction are used. Please note that for privacy purposes, all information is completely anonymous and card numbers are encrypted and cannot be linked to known individuals. A pre-processing was

done on the transaction dataset to make it compatible with the clustering method that has been used. The resulting data file describes the behaviour of a card for a single day: card number, fare type, date, and 24 binary columns for each the hours of the day to indicate if at least a smart card transaction occurred during this hour (0 or 1). We call this observation a "card-day". The resulting datasets contains 4.47 millions "cards-day".

Distance calculation

The k-mean clustering method applied in this study seems classical. However, we defined a special distance function to better capture the similarities between public transport users' journeys.

Consider the dataset in Table 1 that is often used in travel behaviour clustering studies. In this data we have a description of the hourly pattern of public transport use. For example, user #1 was on the transportation network at time H1, user #2 was on the system at time H2, user #8 was there at time H1 and H3, user #9 was on the network from time H1 to time H3. Hence, there is an ordered relation between time intervals: H1 precedes H2, which precedes H3 and so on.

Table 1: Sample dataset for distance calculation example

User #	H1	H2	H3	H4	H5	H6	H7
1	1	0	0	0	0	0	0
2	0	1	0	0	0	0	0
3	0	0	1	0	0	0	0
4	0	0	0	1	0	0	0
5	0	0	0	0	0	1	0
6	0	0	0	0	0	0	1
7	1	1	0	0	0	0	0
8	1	0	1	0	0	0	0
9	1	1	1	0	0	0	0
10	0	0	0	0	0	0	0
11	0	0	0	0	0	0	0
12	1	1	1	1	1	1	1
13	1	1	1	1	1	1	1

In the literature, the similarity of data often relies on some "distance" measures and many metrics do exist (e.g. Euclidean, Manhattan, Hamming) where many variations and weighting of attributes are possible. Besides, these metrics are not well adapted for our purpose because they do not consider relative position of each coordinate.

Table 2: Some distance calculations from the sample dataset

Distance	Euclidean	Manhattan	Hamming
D(1,2)	$\sqrt{2}$	2	2
D(1,3)	$\sqrt{2}$	2	2
D(7,8)	1	1	1
D(7,9)	1	1	1

From a travel behaviour perspective it is obvious that D(1,3) should be greater that D(1,2), which is not the case with regular distance measurement methods. The same applies with D(7,8) and D(7,9). Many others "evidences" for human comprehension do not appear in in the metrics.

Those observations lead us to develop a metric that considers the relative position of the elements in the vector. Our suggested distance is still the Euclidean distance, but calculated on a mapped binary sequence in Cartesian coordinates. The computational and the mathematical advantage of this approach are to be discussed in another article. We suggest to map any binary sequence of any length into a three dimensional Cartesian coordinate system (X,Y,Z) and then use the Euclidean distance on the new three-dimensional space. As a consequence, we gain data reduction of any binary sequence to only three variables. This allows running statistical analyses on huge datasets. Mapping of the binary sequence is implemented in a way to keep the desired distance properties. In the data from Table 1 we map the points such that the Euclidean distance satisfy:

- D(1,3)>D(1,2)
- D(1,2)=D(2,3)
- D(5,6)=D(1,2)
- D(7,8)<D(7,9)
- …

We skip the details of construction of this mapping and just give the heuristic and the formula of our developed methodology. The mapping is easier to understand in the polar coordinate, i.e. in terms of radius, say r, and the angle, say θ.

For a binary sequence, take radius r to be the number of usage n. Note that n is actually the number of ones in the binary sequence. Although, we suggest r to be a more complicated function, being $r = (1 + \frac{1}{n})^n$. The heuristic of using r as a function of n allows mapping binary sequences with the same number of usage on the same circle, and of course sequences with different number of usages on circles with a different radius.

The angle, θ, then, is assigned to the position of the usage (the position of ones). We suggest taking a function that maps a binary sequence of any length, to a value between 0 to 180 (degrees) depending on the position of ones. This mapping allows putting any binary sequence on a half circle.

For instance an appropriate function of angle, gives angle 0 for (1,0,0,0), and gives angle 180 degrees for (0,0,0,1). Transformation from the polar coordinate to the Cartesian coordinate is feasible through the famous transformation X=$r\cos(\theta)$ and Y= $r\sin(\theta)$.

Unfortunately our suggested method, using only X and Y, maps (0,1,1,0) and (1,0,0,1) to the same point in the Cartesian coordinate (X,Y). This is why we define axis Z: in order to differentiate such binary vectors on Z axis.

The pseudo code for calculation of the distance between two binary sequences is as follows:

A) For each binary sequence calculate:

1. The number of ones in the sequence, n

$r = (1 + \frac{1}{n})^n$

2. Find where the ones appears in the vector and take the mean of such positions, call this mean $\overline{p}$, where p is the vector position of ones in the binary vector

3. Calculate the angle $\theta = \frac{(\overline{p}-1)}{(n-1)} \times 180$
4. Find where the one appears in the vector and take the standard deviation of the such positions and dented by s.

(X=$r\sin(\theta)$, Y=$r\cos(\theta)$, Z=s).

B) Calculate the pairwise Euclidean distance on (X,Y,Z), for two sequence

D(1,2)=$\sqrt{(X_1-X_2)^2+(Y_1-Y_2)^2+(Z_1+Z_2)^2}$

The resulting distances for the sample of 13 users presented in Table 1 can be found in Table 3. One may note that the results are more suitable to a travel behaviour analysis.

Table 3: Resulting distance matrix between the 13 users

	1	2	3	4	5	6	7	8	9	10	11	12	13
1	0	41	79	111	152	157	29	55	52	79	79	137	137
2	41	0	41	79	136	152	29	35	28	79	79	104	104
3	79	41	0	41	111	136	67	55	52	79	79	72	72
4	111	79	41	0	79	111	103	90	90	79	79	55	55
5	152	136	111	79	0	41	155	148	151	79	79	104	104
6	157	152	136	111	41	0	166	165	167	79	79	137	137
7	29	29	67	103	155	166	0	28	25	90	90	120	120
8	55	35	55	90	148	165	28	0	11	94	94	96	96
9	52	28	52	90	151	167	25	11	0	96	96	101	101
10	79	79	79	79	79	79	90	94	96	0	0	112	112
11	79	79	79	79	79	79	90	94	96	0	0	112	112
12	137	104	72	55	104	137	120	96	101	112	112	0	0
13	137	104	72	55	104	137	120	96	101	112	112	0	0

Data clustering

As the metric in Cartesian coordinate is Euclidean, it is straightforward to implement clustering methods developed for Euclidean distance, such as the famous k-means algorithm. As the number of variables is only three, (X,Y,Z), the k-means algorithm is fast to run on our massive data (4.47 million observations). Furthermore, the Cartesian heuristic allows even visualizing the data, seeing for instance Figure 1, where we show 4.5 million points. Each circle shows a mapped binary sequence of size 24 (hours). The size of circles is proportional to the frequency of that Cartesian point. Figure 1 clearly suggests three clusters, so we run the k-means algorithm with three centres k=3 on the Cartesian data. The centres are shown using filled circles. The k-means algorithm run on 4.47 million observations took only 4 seconds using R statistical software (http://r-project.org) on a common desktop computer due to the effective dimensionality reduction.

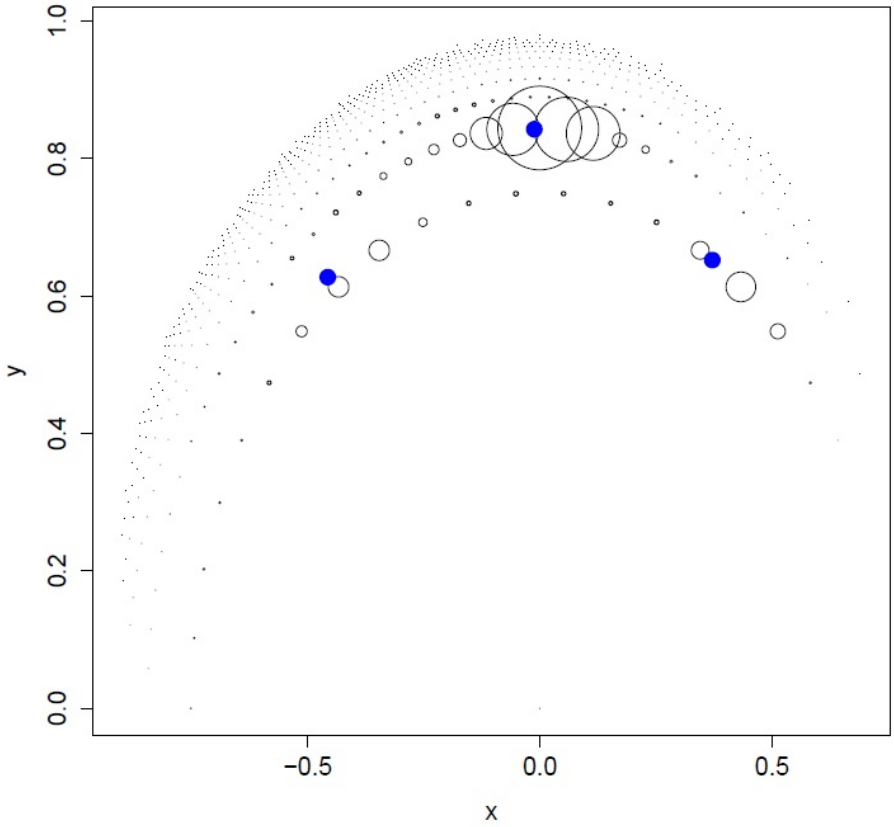

Figure 1: Dataset mapping

RESULTS

The results presented here are based on the clustering analysis of the 4.47 millions cards-day, as described before.

Clusters description

Figure 2 presents the distribution of the average values of each cluster's observations. It reveals that cluster C1 is characterized by a two-peak distribution of the transactions during a typical day of travel. This is typical of pendulum AM and PM peak period travel. The second cluster (C2) is related to morning travels. A strong AM peak mostly characterizes it and it has many transactions reported from 9:00 AM to 12:00 PM. At last, there are very few transactions in the afternoon for this cluster. In cluster C3, trips are made mostly at PM peak and during the evening. Note that the PM peak is weaker than for cluster C1 and that travel is more spread during the day. These are average values: it does not mean than a user in cluster C2 do not make any trips at the PM peak; however, it means that the behaviour of this user is closer to the behaviour of other C2 people than the behaviour of any people in other clusters.

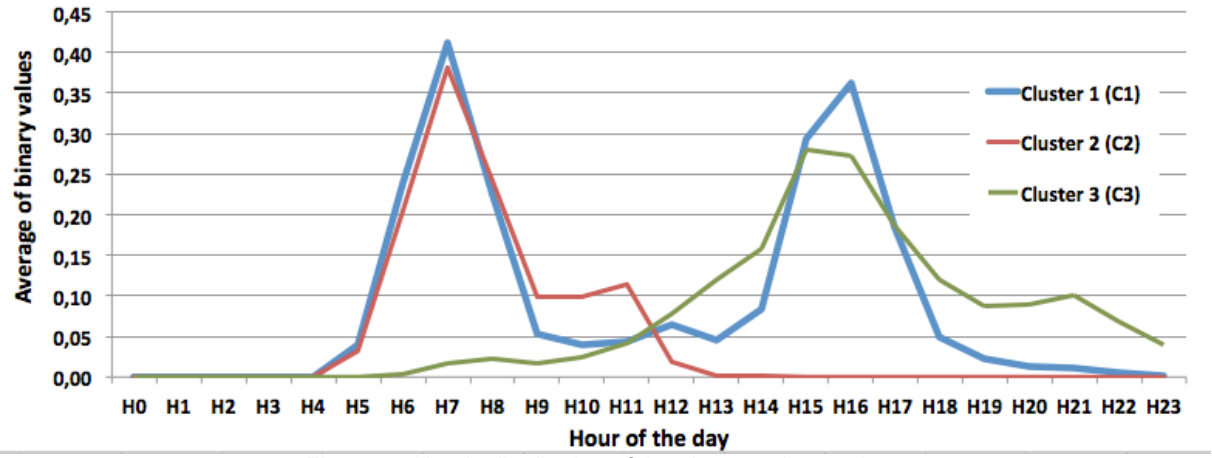

Figure 2: Hourly distribution of the clusters observations

Day of week

Let us look at the proportion of cards-day associated to each cluster, by day of the week (Figure 3). It shows that during working days (from Monday to Friday), the proportion of C1 cluster (pendulum AM-PM trips) is much higher than the other clusters. However, on Friday, the proportion of C2 and C3 clusters are higher, suggesting that the travel behaviour of public transport users is a little bit different on this day. On weekends, the C3 cluster dominates the others. People move later in the afternoon, and the trips are less characterized by pendulum movements like in working days.

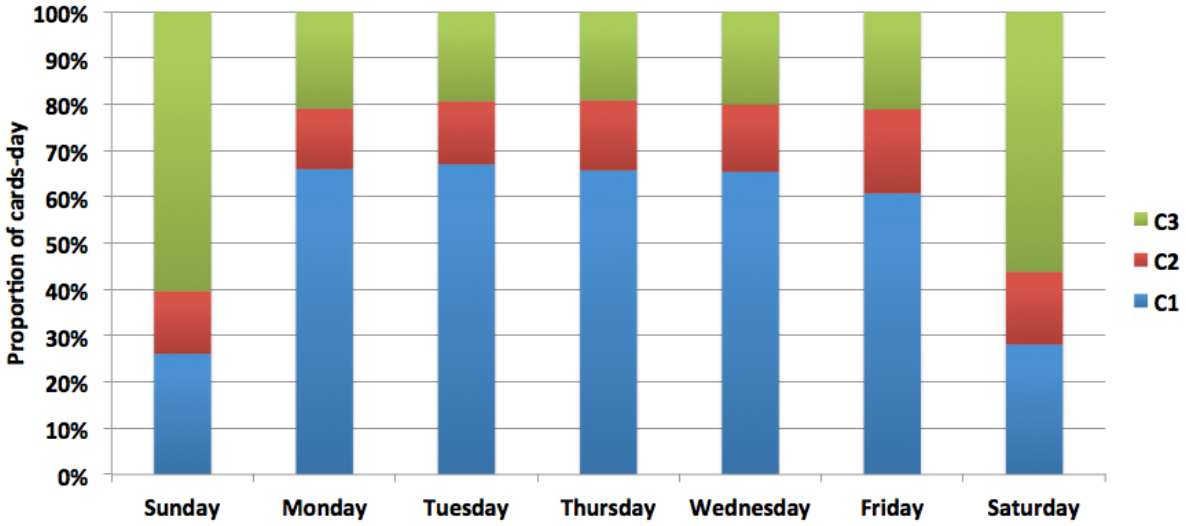

Figure 3: Distribution of the cards-day belonging to cluster, by day of week

Fare type

It is interesting to look at the distribution of the cluster by fare type. Fare type characterizes both the age of the user (may be student, adult, senior), the type of service used (express routes, longer interzone routes, regular routes) and the type of payment (paid at booth or by pre-authorized banking payment).

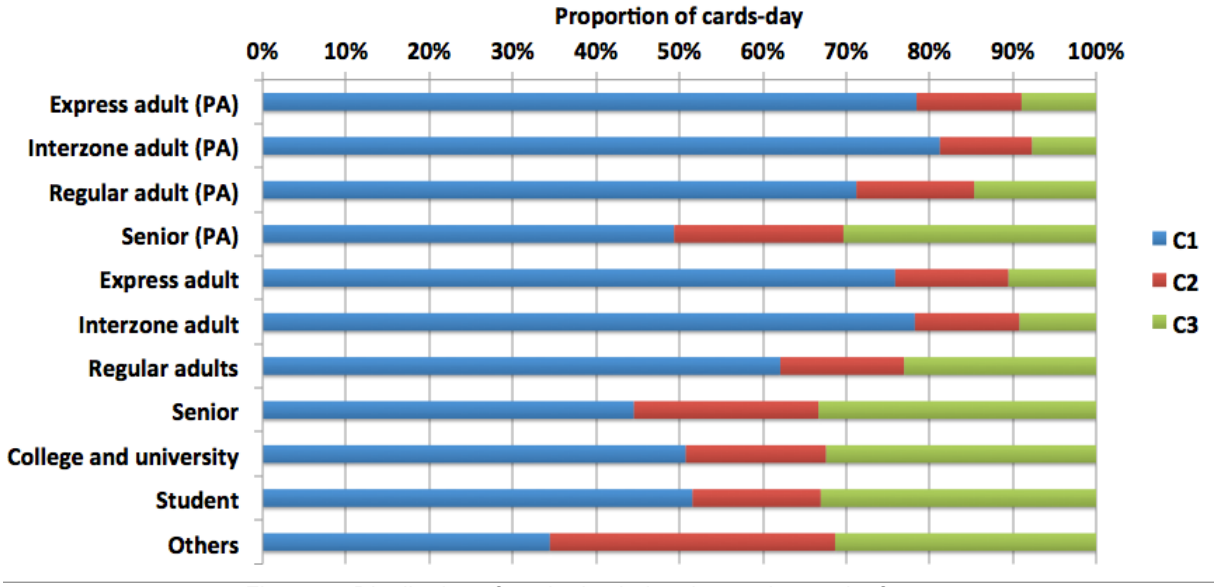

Figure 4: Distribution of cards-day belonging to clusters by fare type

As seen before, we can associate the C1 cluster to a loyal, regular pendulum movement AM and PM. It seems likely that cardholders that use pre-authorized payment (PA) are more loyal to the system than their counterparts. Express and interzone adult fare holders board routes that travel longer distances and in a fastest way. The cards-day of these fares belongs up to 80% to the C1 cluster. Senior users are characterized by a higher rate of C2 and C3 clusters, suggesting they are less commuting than others. Students also have a higher belonging to C2 and C3 cluster. This may be related to the irregular schedules of college and university students, and to the earlier end of school days, compared to working days. The "others" fare type is associated to special events; it is not surprising that the behaviour is more mixed than for other groups.

Dominant cluster

At last, we present an analysis of the dominant clusters of each card. To identify the dominant cluster, we analyse the sequence of cards-day for each card and we retain the cluster in which the card was the most often found. For example, if a card is found in C1 65% of the time, 30% of the time in C2 and the remaining in C3, the dominant cluster for this card will be C1. Dominant clusters analysis is important for public transport planners because it shows if a given user (or his card) is loyal to his behaviour or not. Figure 5 shows the distribution of cards given the part of the dominant cluster in their set of cards-day. The analysis is done only for working days (Monday to Friday). A dominant found between 30% and 40% means a weak dominance because the other two clusters are almost equal in proportion to the dominant one. On another hand, a part of 100% means that the cluster is the only one found for this card. Knowing that the analysis is for an entire year, this tells us that the user followed the same behaviour for all his trips during a long period.

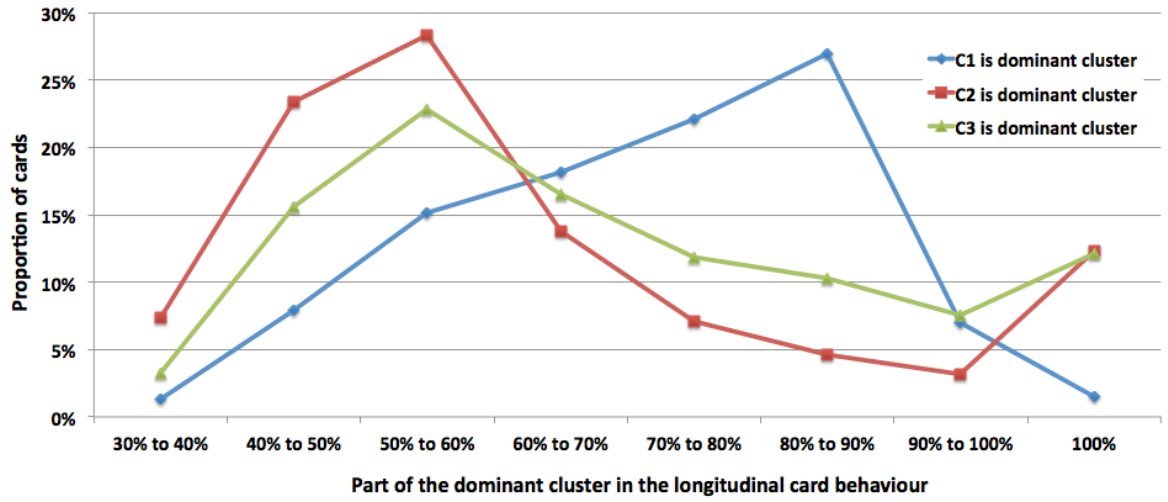

Figure 5: Distribution of the relative importance of dominant clusters related to each card (Mon. to Fri. only)

The figure shows that cluster C1 is characterized by a higher loyalty, having more cards located in the right part of the chart. The C3 cluster has a more dispersed distribution, with a peak at the 50% to 60% class. However, it has a higher 100% share than the C1, showing that some users always behave the same during the period. The C2 cluster distribution is located mostly at the left, suggesting that these cardholders are more likely to change their behaviour during a long period.

CONCLUSION

In this paper, we presented an analysis of public transport smart card transactions with the help of k-means data mining technique. One year of data gathered from the *Société de transport de l'Outaouais* was pre-processed in 4.47 millions cards-day that were the unit of analysis for the k-means. However, instead of using traditional distance between observations, we proposed an innovative distance calculation method that takes into account the location of the observations in the 24-hour binary vectors that we used to characterize each card-day.

Three clusters were obtained from the k-means analysis. Cluster C1 is characterize by pendulum AM-PM travel. It is the most used in the behaviour of the adults, however less used by seniors and students. Cluster C2, seen as morning trip, occupies a smaller part in dominant clusters. Cluster C3, related to PM and evening trips, is mostly seen in weekends.

Numerous perspectives arise from this work. First, there is a need to provide a mathematical proof that the distance method we propose is better than the existing ones. Second, the k-means analysis is to be improved and applied to smaller subsets; it may find more clusters or other set of clusters according to different attributes. Third, the technique can be applied to other sorts of vectors, not only including transaction times, but also the location of boarding on the territory, the route sequences, route types, etc.

ACKNOWLEDGEMENTS

The authors wish to acknowledge the support of the *Société de transport de l'Outaouais*, who provided the data for this study.

REFERENCES

Anand, S.S., Büchner, A.G. (1998) Decision Support Using Data Mining, Financial Times Pitman Publishers, London, UK.

Bagchi, M., White, P.R., 2005. The potential of public transport smart card data. Transport Policy 12, 464–474.

Chapleau, R., Chu, K.K., 2007. Modeling transit travel patterns from location-stamped smart card data using a disaggregate approach. In: 11th World Conference on Transportation Research, Berkeley, California (CD-ROM).

Chu, K.K., Chapleau, R. (2008). Enriching Archived Smart Card Transaction Data for Transit Demand Modeling. Transportation Research Record: Journal of the Transportation Research Board, No. 2063, Transportation Research Board of the National Academies, Washington, DC, pp. 63–72.

Devillaine Flavio, Munizaga Marcela, Trépanier Martin (2012), Detection of Activities of Public Transport Users by Analyzing Smart Card Data, Transportation Research Record: Journal of the Transportation Research Board, no. 2276, vol. 3, pp. 48-55.

Gordon, J.B., Koutsopoulos, H.N., Wilson, N.H.M., Attanucci, J.P. (2013). Automated Inference of Linked Transit Journeys in London Using Fare-Transaction and Vehicle-Location Data, In: Presented at the 92nd meeting of the Transportation Research Board, paper # 13-0740, Washington, D.C.

Ma, X., Wang, Y., Chen, F., Liu, J. (2013a). Transit Smart Card Data Mining for Passenger Origin Information Extraction. In: Presented at the 92nd meeting of the Transportation Research Board, paper # 13-2156, Washington, D.C.

Ma, X., Wu, Y.-J., Wang, Y., Chen, F., Liu, J. (2013b). Mining Smart Card Data for Transit Riders' Travel Patterns. In: Presented at the 92nd meeting of the Transportation Research Board, paper # 13-3460, Washington, D.C.

Morency, C., Trépanier, M., Agard, B. (2006). Analysing the variability of transit users behaviour with smart card data. In: The 9th International IEEE Conference on Intelligent Transportation Systems – ITSC 2006, Toronto, Canada, September 17–20.

Morency, C., Trépanier, M., Agard, B. (2007). Measuring transit use variability with smart-card data. Transport Policy 14 (3), 193–203.

Munizaga, M., Palma, C., Mora, P. (2010). Public transport OD matrix estimation from smart card payment system data. In: Presented at the 12th World Conference on Transport Research, Lisbon, Paper No. 2988.

N. Lathia, C. Smith, J. Froehlich, L. Capra, Individuals among commuters: Building personalised transport information services from fare collection systems, Pervasive and Mobile Computing, doi: 10.1016/j.pmcj.2012.10.007, 2012.

Park, J.Y., Kim, D.J. (2008). The Potential of Using the Smart Card Data to Define the Use of Public Transit in Seoul. Transportation Research Record: Journal of the

Transportation Research Board, No. 2063, Transportation Research Board of the National Academies, Washington, DC, pp. 3–9.

Pelletier, M.-P., Trépanier, M., Morency, C. (2011). Smart card data use in public transit: A literature review. Transportation Research Part C: Methodological, 19, 557-568.

R. Xu, and D. Wunsch II, Survey of Clustering Algorithms, IEEE Trans. On Neural networks, Vol. 16, No. 3, 2005.

Reddy, A., Lu, A., Kumar, S., Bashmakov, V., Rudenko, S., 2009. Application of Entry-Only Automated Fare Collection (AFC) System Data to Infer Ridership, Rider Destinations, Unlinked Trips, and Passenger Miles. 88th Annual Meeting of the Transportation Research Board, Washington, 21 p.

Seaborn, C., Wilson, N.H., Attanucci, J. (2009). Using Smart Card Fare Payment Data to Analyze Multi-Modal Public Transport Journeys (London, UK). 88th Annual Meeting of the Transportation Research Board, Washington, 16 p. (CD-ROM).

Tibshirani, R., Walther, G. and Hastie, T. (2001). Estimating the number of data clusters via the Gap statistic. Journal of the Royal Statistical Society B, 63, 411–423.

Trépanier, M., Chapleau, R., Tranchant, N. (2007). Individual trip destination estimation in transit smart card automated fare collection system. Journal of Intelligent Transportation Systems: Technology, Planning, and Operations 11 (1), 1–15 (Taylor & Francis).

Trépanier, M., Morency, C., Agard, B., 2009. Calculation of transit performance measures using smartcard data. Journal of Public Transportation 12 (1), 79–96.

Trépanier, M., Nurul Habib, K.M., Morency, C. (2012). Are transit users loyal? Revelations from a hazard model based on smart card data. Canadian Journal of Civil Engineering, 39, 610-618.

Zhao, J., A. Rahbee, and N. H. Wilson (2007). Estimating a Rail Passenger Trip Origin-Destination Matrix Using Automatic Data Collection Systems. Computer-Aided Civil and Infrastructure Engineering. Vol. 22(5), pp. 376-387.