Leveraging system intelligence from massive smart card database to support operational and strategic objectives of a transit agency

Ka Kee Alfred Chu, Ph.D., Expert-modeling, Agence métropolitaine de transport (AMT)

Paper prepared for presentation

at the Best Practices in (Urban) Transportation Planning (A) Session

of the 2014 Conference of the

Transportation Association of Canada

Montréal, Québec

ABSTRACT

Transit agencies have long been operating in data-poor environments. They have been relying on labour-intensive and project-specific data collection, which are often sparse and costly. Recently, the adoption of technologies that generate passive data streams has become increasingly common. One of such technologies is the smart card automatic fare collection (AFC) system which provides a massive database containing detailed and up-to-date fare validation records. An important preoccupation of both practitioners and researchers is to transform the raw and often partial data into useful information.

Various aspects of a transit agency, from day-to-day operations to strategic rethinking of the transit system, can benefit from such intelligence. While data timeliness of smart card data is an obvious advantage over other data collection methods, data organization and processing are important issues. Each application, aiming to tackle a specific problem, requires a different approach and data enrichment procedure. Drawing data and experience from the Montréal region, which currently operates an extensive multi-region, multi-modal and multi-operator smart card AFC system that generates two million transactions per day, this paper presents contributions of smart card intelligence within a regional transit authority context by showcasing several real-world applications on transit services, transit equipment and fare use.

Smart card data analytics provide different departments within the organization and among partner organizations in the region access to up-to-date and previously unavailable information. This in turn allows them to make more informed decision on policies and plans. The future of public transit relies greatly on adequate data and intelligence. The next steps would be to expand the number of applications, to put in place a structure that facilitates systematic and corporate use of the data and to increase cooperation among partner organizations.

INTRODUCTION

CONTEXT

Transit agencies have long been operating in data-poor environments. They have been relying on labour-intensive and project-specific data collection, which are often sparse and costly. Recently, the adoption of technologies that generate passive data streams has become increasingly common. One of such technologies is the smart card automatic fare collection (AFC) system.

THE OPUS AFC SYSTEM

Montreal is one of the first major metropolitan areas to adopt smart card AFC system in North America. The system, named OPUS, is multi-region, multi-modal and multi-operator. The system started its operation in 2008 and now includes 20 operators in the Montréal and Québec City regions. It is used in the commuter rail, metro, bus, ferry and other transit equipment. The AFC system accepts primarily reloadable and disposable smart cards. Some other methods of payments such as magnetic cards and cash are partially integrated into the system.

DATA GENERATED

A transaction record is generated whenever a fare product is validated by an equipment. In an average weekday, there are about two million fare validations. This provides the basis for a massive database containing precise and up-to-date details on the activities within the transit system. The database has now accumulated more than 2 billion records since 2010. An important preoccupation for both practitioners and researchers is to transform the raw and often partial data into useful information. This paper showcases how to derive business intelligence from the smart card fare validation database within a regional transit agency context. The data and applications are based on the Montréal region and the works by AMT – Agence métropolitaine de transport. Various aspects of a transit agency, from day-to-day operations to strategic rethinking of the transit system, can benefit from such intelligence. While data timeliness of smart card data is an obvious advantage over other data collection methods, data organization and processing are important issues. Each application, aiming to tackle a specific problem, requires a different approach and data enrichment procedure. But in general, it involves:

- Storing the data in manageable units
- Preparing reference data tables
- Enriching the data by
  - Joining reference data
  - Deriving additional variables
  - Merging external datasets
- Summarizing the data
- Presenting the results

DATA AND APPLICATIONS

The database is comprised of archived fare validation records. The data are multi-dimensional. Each record contains:

- Spatial and service component
  - Fare validation location: train and metro station (fixed)
  - Fare validation location: bus (variable), determinable with GPS if equipped
  - Route, direction, run of the service (if applicable)
  - Fare zone of the vehicle
- Temporal component
  - Fare validation time and date to the nearest second
  - Type de fare validation: first boarding, transfer boarding
- Card component
  - Fare medium
  - Anonymized card number
- Fare component
  - Debited fare product
  - Period of validity of the fare product
  - Anonymized fare contract number

Since the smart card AFC is an entry-only system, no data on alighting is collected. Sale data of fare products and fare control are also generated by the system and can be analyzed alongside fare validation data. Applications that will be presented can be grouped into three main categories:

- Transit service and equipment
- Fare and marketing
- Planning and modelling

TRANSIT SERVICE AND EQUIPMENT

Express route Chevrier

The most basic indicator that smart card data can provide relates to ridership. It can be used to follow long-term trend, monitor specific events and produce more in-depth diagnostics. Figure 1 shows daily ridership of express route Chevrier that operates between Brossard in the South Shore and downtown Montréal via the Champlain Bridge. The data clearly shows distinctive weekly and seasonal ridership patterns. In addition, specific events can be identified and analyzed. The systematic collection, the timeliness and the level of details of the data are only a few advantages over traditional ridership count methods.

The typical weekly pattern is exemplified by the high level of ridership during weekdays and the low level of ridership over the weekend as the route is primarily designed for commuting. The level of ridership over the year can also be divided into four distinctive seasons: spring, summer, fall and the holiday seasons. Ridership is the highest and the most constant during the fall season. Ridership level in spring is also high but it tails off towards summer as students progressively enter summer holidays. Summer ridership is significantly lower as both workers and students are having their summer holidays. During the two-week holiday season, the ridership is low as many work and study trips turn to a halt.

Other specific events are also captured by smart card data. For example:

- A (referring to Figure 1): spring break for university students where a significant drop in ridership during a week can be seen
- B: a 25cm snow storm in Montreal affecting the ridership for the day
- C: long weekends in the last week of June (Québec's national holiday) and first week of July (Canada day) affect not only the Mondays which are statutory holidays but the entire two weeks as some workers take additional days off
- D: construction holidays (two weeks) affect ridership as members from the same household take days off at the same time

The ridership information can be used for short- and medium-term planning in order to optimize transit fleet and service, as well as other related duties such as staffs responsible for fare control and safety.

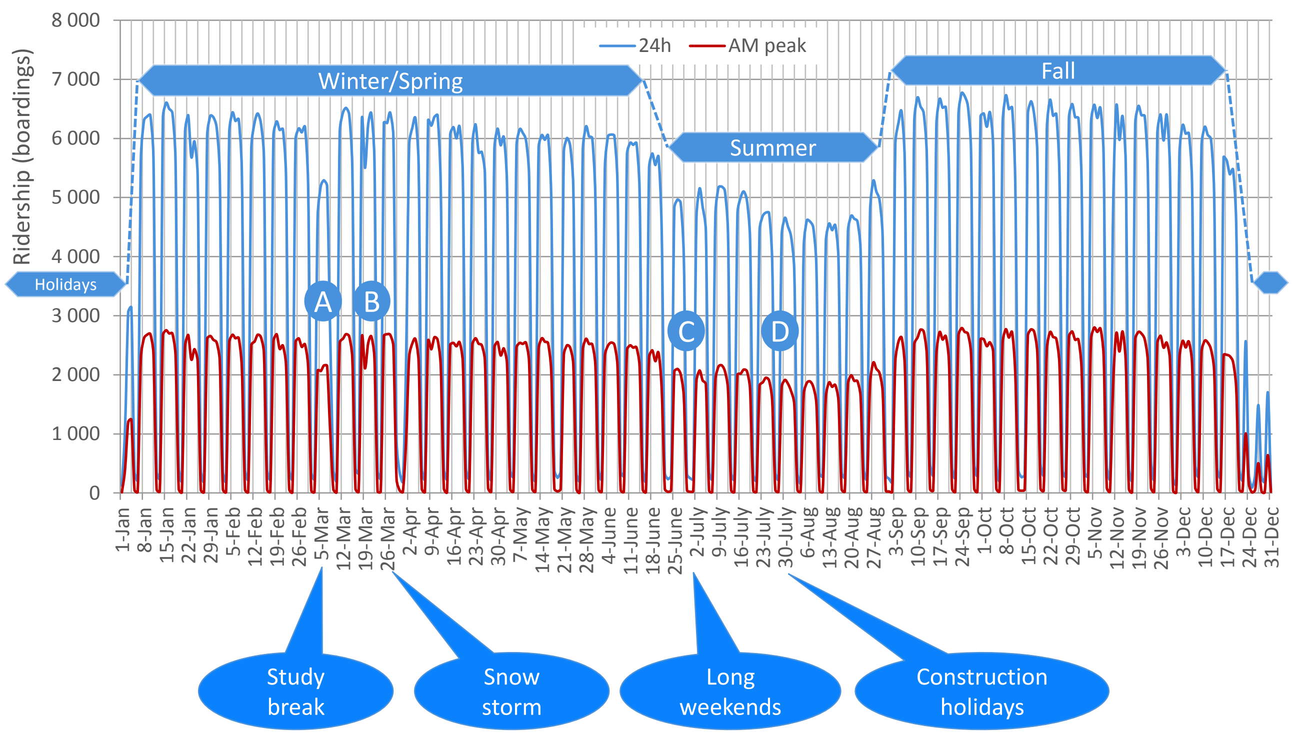

Figure 1 Ridership statistics of the express route Chevrier

Smart card data also contributes to the study of daily travel pattern and to questions such as the relevance of peak and off-peak fare measure by providing ridership at a very fine temporal level. Figure 2 shows the number of boardings on express route Chevrier per 15-minute period in a typical weekday. They are grouped into four main periods:

- AM peak (6:00 – 8:59): 41% of daily ridership
- Midday (9:00 – 15:29): 12% of daily ridership
- PM peak (15:30 – 18:29): 36% of daily ridership
- Evening (18:30 – 24:59): 11% of daily ridership

As a commuter route, boardings in the AM and PM peaks account for the majority (77%) of total daily ridership, with few users travelling during midday and evening. The figure also reveals that the AM peak is more intense while the PM peak is more spread out and has several “sub-peaks”. The information can be used for service planning and fare analysis in which the level of comfort can be a contributing factor.

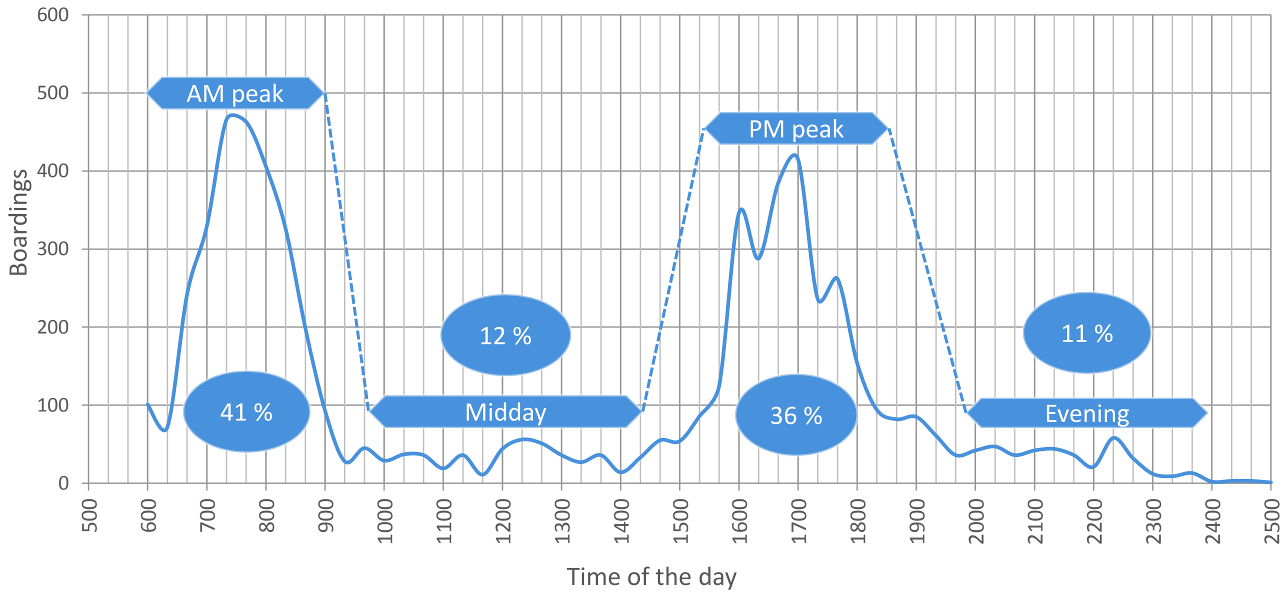

Figure 2 Temporal distribution of boardings of the express route Chevrier in a typical day

In addition, the travel pattern of individual cardholders of the express route Chevrier is examined. This provides a more in-depth understanding on commuter’s needs and helps transit authority to plan more suitable service. Figure 3 is an aggregate temporal travel matrix showing the departure time and time of the return trip on the same route at the individual level for a typical day. About 24% of the users travel one-way on this route and probably return on another route or by another transport mean. Almost 50% of the users depart during the AM peak and return during the PM peak, a pattern typical of a commuter.

	Return trip						
Departure	0:00 to 5:59	6:00 to 8:59	9:00 to 11:59	12:00 to 15:29	15:30 to 18:29	18:30 to 23:59	One-way
0:00 to 5:59	0%	0%	0%	0%	1%	0%	0%
6:00 to 8:59		0%	1%	5%	49%	8%	8%
9:00 to 11:59			0%	1%	4%	3%	2%
12:00 to 15:29				0%	1%	2%	3%
15:30 to 18:29					0%	1%	7%
18:30 to 23:59						0%	4%

Figure 3 Temporal travel matrix of the express route Chevrier in a typical day

Subway system

The subway system, known as the Métro, is an integral part of the transit network in the Montréal region. The stations are mostly located on the island of Montréal and the network is connected to Laval in the North Shore and to Longueuil in the South Shore. The ridership is about 1.2 million in an average weekday. Smart card data provide detailed information on the network, its stations and its users. Figure 4 shows the total number of entries at each station in an average weekday. The ridership is proportional to the size of the pie. Meanwhile, the colour indicates the proportion of ridership with local STM fare or integrated TRAM fare products. The first observation is that ridership and fare varies greatly from one station to another. The second observation is that variations are consistent with the network setting. Stations located in the central business district and the terminal stations usually have a higher ridership. Intermodal stations, termini and special trip generators such as universities and hospitals, usually have a higher proportion of integrated fare users.

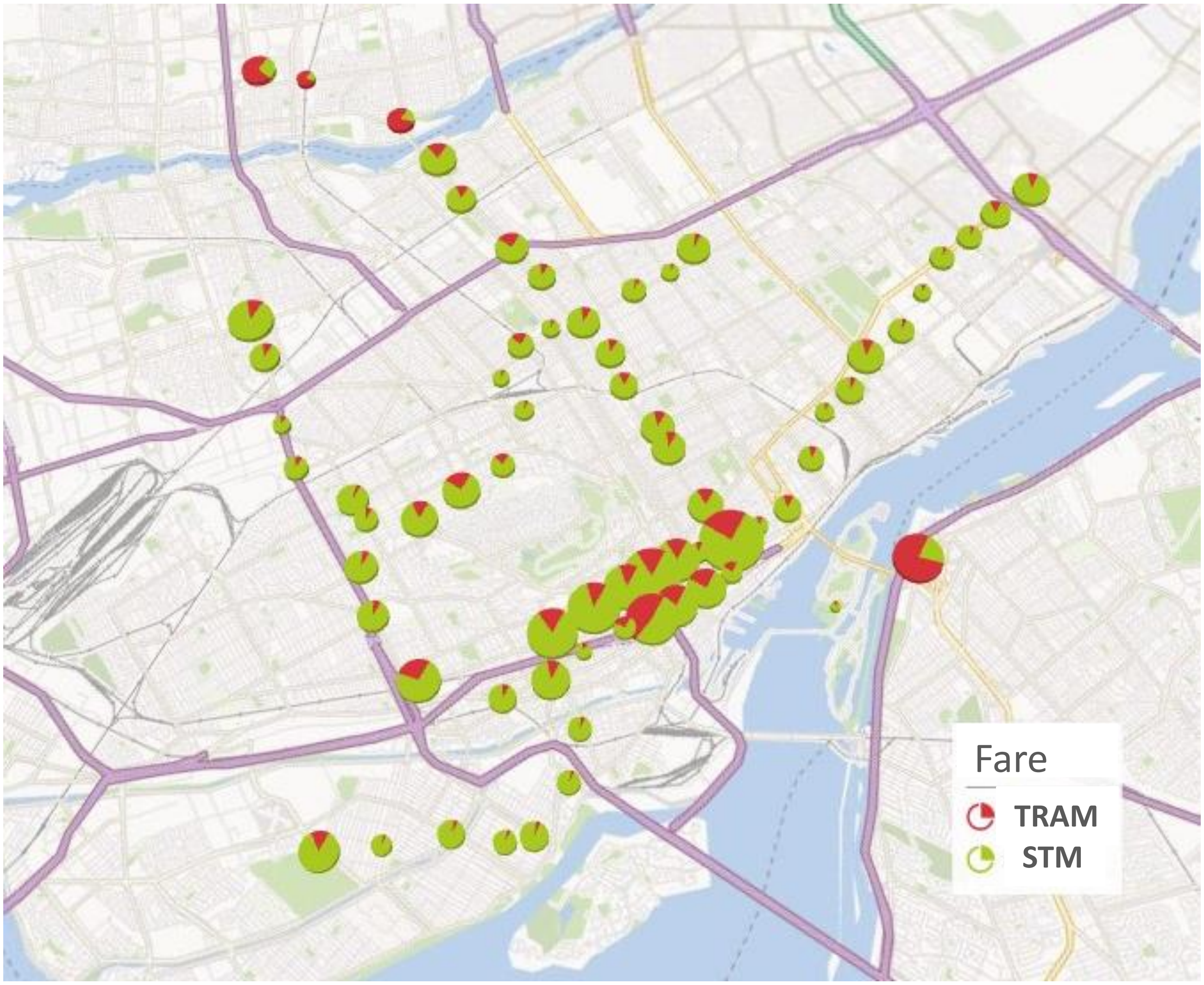

Figure 4 Ridership and fare use of the subway network in a typical day

Vélostation

Some transit equipment, at least in the form of pilot project, uses the OPUS card to control access. Vélostation, literally "bike station" in English, is an enclosed space with racks dedicated for bicycle storage. It encourages intermodal transfer between bike and train at a location where park-and-ride facility is saturated. As a pilot project, some clients with annual subscription are given the opportunity to test the facility. A fare product on the OPUS card allows the opening of the door into the enclosed area.

Similar to transit service, the usage pattern of this transit equipment is analyzed. The daily number of entrances into Vélostation can be compiled. However, the maximum number of occupied spaces is a more relevant indicator and can be derived using some simple assumptions. The main one is that users validate their smart card as least once when they enter Vélostation (either to deposit or to take out their bikes). The first validation of the card is considered as bike deposit and the last validation of the day is considered take out. Some may follow other users into Vélostation without using their card. This may be especially true when users return on the same train in the evening. For this reason, a space is considered occupied until mid-night if there is no second validation on the same card on the same day. The results are validated with on-site bike counts.

Figure 5 explores the usage pattern over the month. As a transit equipment located on a commuter rail line, it is expected that usage is tightly linked to the work week. The figure shows that usage on Saturdays and Sundays is low, as well as on Thanksgiving Monday. Although no statistical relationship is established, the change in temperature and precipitation in this month seems to have a slight effect on usage. There seems to be fewer users on rainy days. As the average weekly temperature decreases, so does the usage.

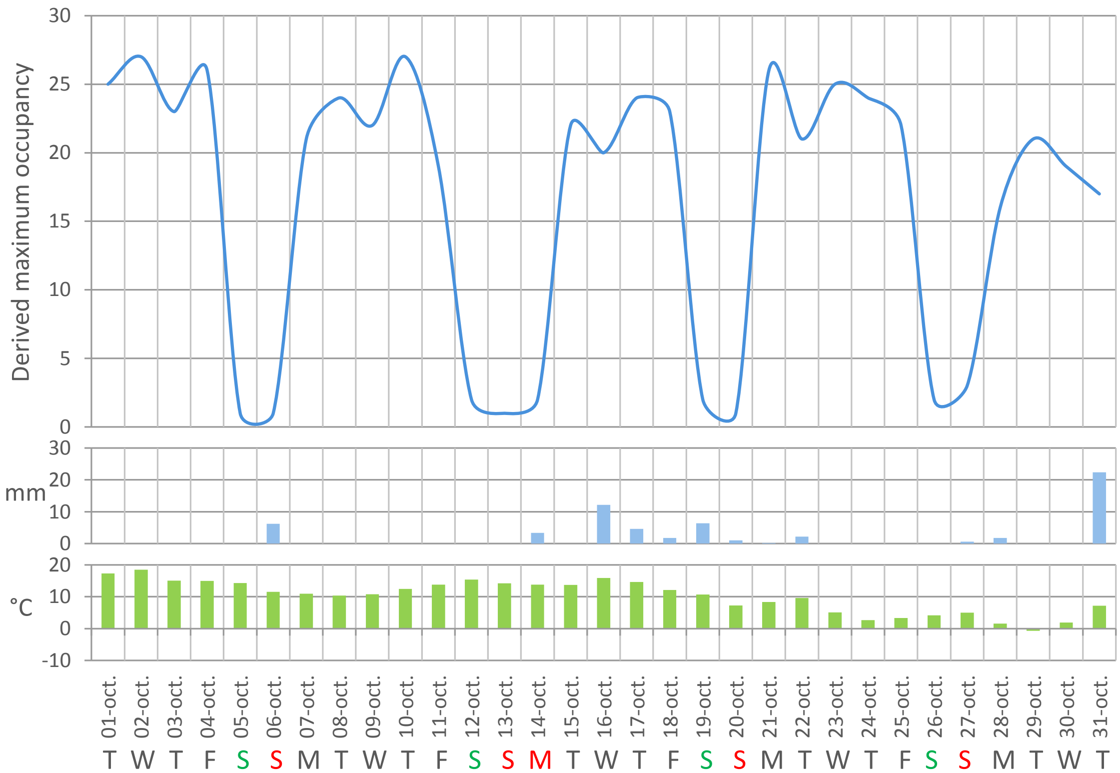

Figure 5 Maximum occupancy of Vélostation and weather over a month

The timely information is very useful for a pilot project with a lot of novelties and questions unanswered. From a practical point of view, information can be used to optimize the number of cardholders permitted to use the facility.

FARE AND MARKETING

As an agency in charge of regional transit planning and coordination, AMT is responsible for issuing integrated fare products and fare revenue sharing. Other than regular fare products, special fare products and fare promotion are managed by AMT to encourage the use of public transit. Analyses with smart card data are crucial to decision making during the planning, execution and the follow-up stages of those programs.

Planning, execution and follow-up of regional fare promotions

During the planning stage, numerous assumptions have to be made based on limited data and experience. The decisions in this stage often have significant impacts on the costs and the prospect of achieving the desired objectives. Smart card data are used to establish assumptions and scenarios in order to help decision making by AMT and its partners.

During the execution stage, analyses by smart card data allow the monitoring the progression of the program as well as the severity of the situation the program is designed to tackle. Figure 6 shows the number of validated complimentary tickets over a month. The government, AMT and its partner want to alleviate traffic congestion on the Champlain Bridge that undergoes urgent road work. Complimentary tickets are distributed to the public at strategic locations in order to promote modal shift towards public transit. The analysis shows the number of validated promotional tickets by date. A substantial increase in the use of the complimentary tickets is noticed during the urgent road work when compared to the baseline use of previously distributed tickets. The feedbacks from smart card analysis allow AMT and its partners to evaluate and adjust mitigation measures in a timely manner.

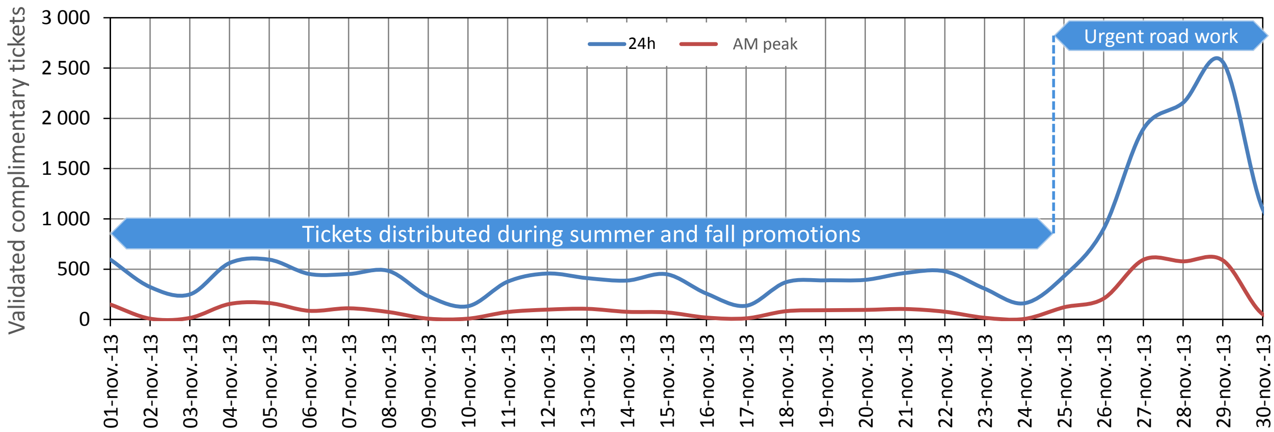

Figure 6 Use of complimentary ticket over a month

During the follow-up stage, revenue and cost sharing is done based on the number of tickets validated within each transit network as opposed to the number of tickets issued. This allows simpler operations and a fairer revenue distribution or compensation mechanism. Data processing and enrichment procedures are designed to produce specific usage indicators from smart card data. Most importantly, the knowledge and intelligence produced by smart card data would accumulate and feed future initiatives.

Regional fare revenue sharing

The integrated fare products generate annually more than 150 million dollars in passenger revenue which is shared among transit operators. A mechanism, based on travel survey and transit model, has been in place before the smart card system. A roadmap has been put in place in order to integrate smart card data into the mechanism.

User characterization

Ridership is also analyzed from the user perspective. Although users cannot be identified, some characteristics of the route are revealed by the choice of fare products. Figure 7 shows the proportion of monthly ridership of express route Chevrier by fare category, fare package and fare type. They are derived from the fare product in the fare validation record. Cash fare represents about 1% of ridership and is not included in the figure.

- Fare category: the majority of ridership is associated with regular fare, which is consistent with the commuting nature of the route. Full-time university and college students can fall into both the “student” and “reduced” categories. Their number is small compared to regular fare users.
- Fare package: the majority of ridership comes from holders of monthly pass or annual subscription which allow unlimited travel. This is consistent with the significant level of rebate given to those two packages with respect to trip-based fare.

- Fare type: local fare only allows travel within the same operator and integrated fare allows interchange among operators. The observation reveals that the use of the subway, which requires an integrated fare or a separate fare, is important but not prevalent.

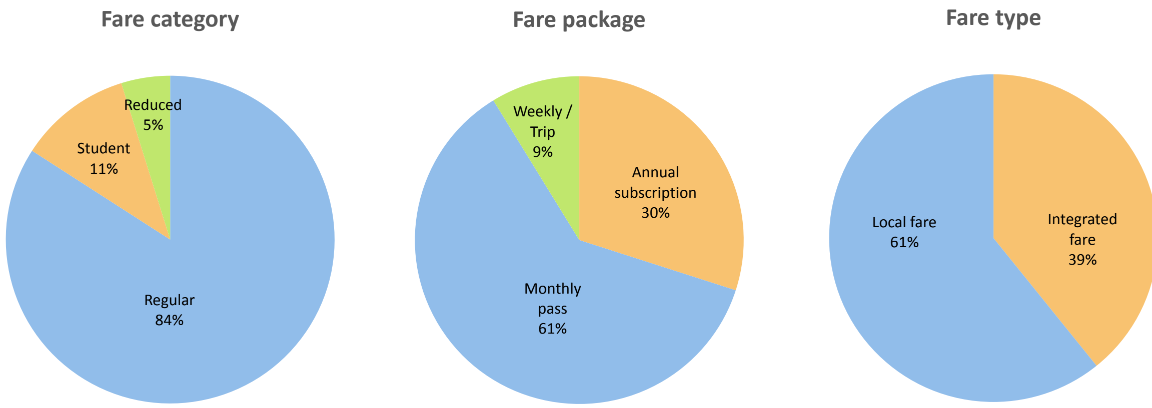

Figure 7 Fare use of the express route Chevrier

Another way to look at ridership is to look at use frequency at the individual level. Although Figure 7 gives an overview of the composition of fare products on the route, it does not reveal how often each cardholder uses the route and how diverse the users are. Figure 8 draws a different portrait. It screens all the OPUS reloadable cards transiting on express route Chevrier and counts the number of boardings made on the route within a month. The cards are then grouped into three classes, namely "loyal", "typical" and "potential" users, according to the frequency of use. It reveals that about 80% of the ridership comes from about 40% of cardholders while the rest of the ridership comes from a very diverse pool of cards. Although it can sometime be explained by operational constraints of the AFC system (such as lost or expired card replacement), this longitudinal analysis of card use provides an extra dimension on the clientele of the route.

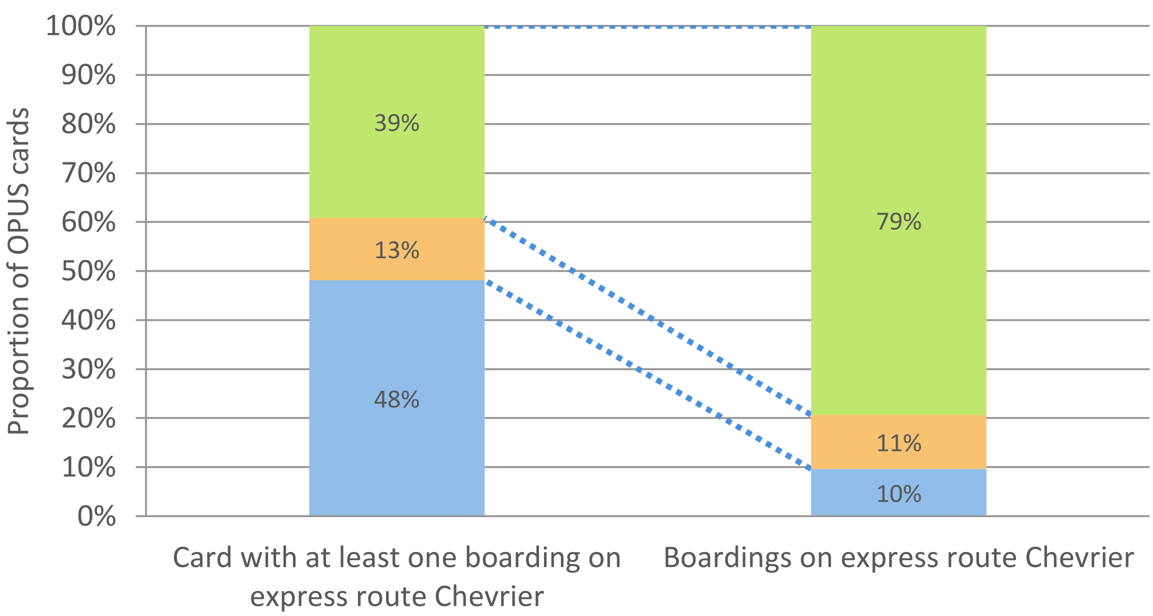

Figure 8 Contribution of ridership of the express route Chevrier from three classes of cardholders

Fare purchase and use

One example is to examine whether to renew the program that offers additional rebate to annual subscription. The program has been popular among transit users but there seems to be signs that the market has been saturated. Assuming the program would be attractive to cardholders who buy 10 or more monthly passes over the year, the goal is to estimate the remaining size of this pool of cardholders. Fare products used by each individual card over the past 12 months can be derived from smart card data, thus providing a good estimation to the size of the pool.

Long-term regional fare reform

A strategic planning on regional fare reform is in process. It aims to simplify the fare structure for transit users. Smart card AFC system provides regional and systematic sale and consumption information on each fare product for analyses and discussions.

TRANSIT PLANNING AND MODELLING

Project planning

Subway fare validations contain entrance location but not alighting location. The partial knowledge can be enriched by merging data from travel survey. The latter have numerous subway itineraries with origin and destination stations, line taken and transfer station. Figure 9 shows the load profiles of Montmorency-bound orange line over the three-hour AM peak period. The maximal load point is located between Place St-Henri and Lionel-Groulx stations. One profile is produced for each weekday and the load on Friday is significantly lighter than the rest of the week. Such analysis provides insights on current state of the system and guidance for future investment.

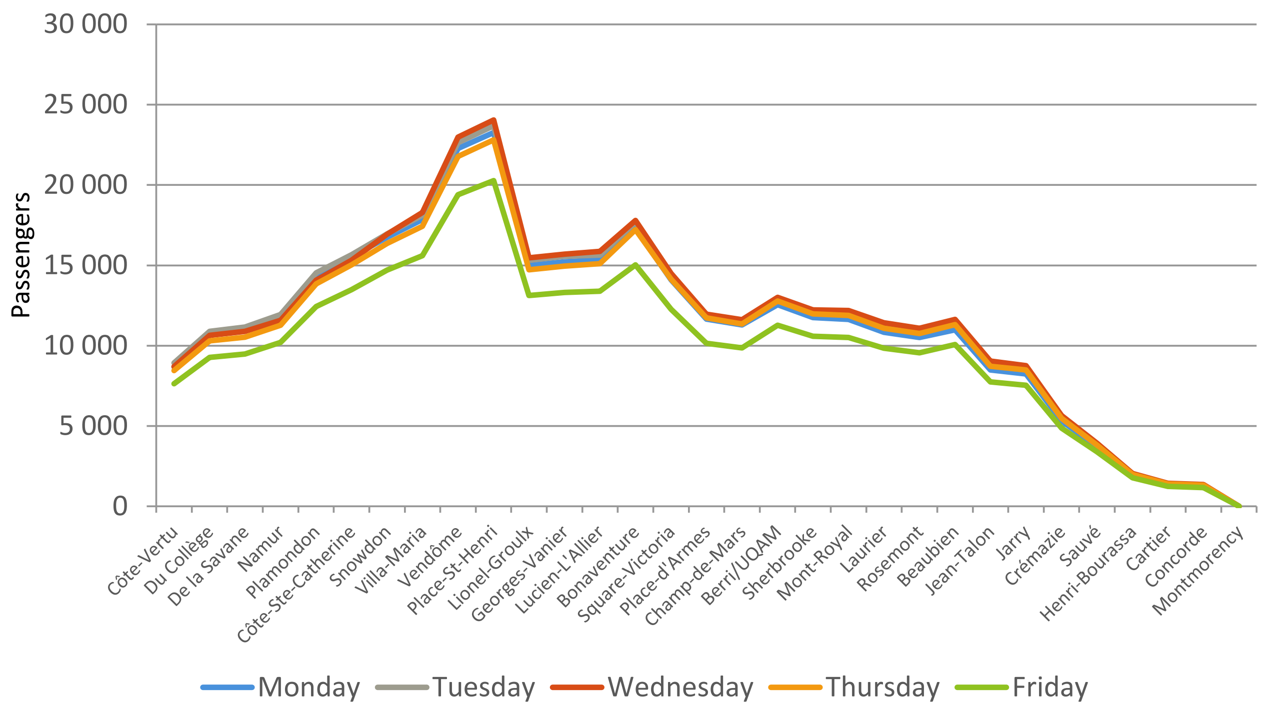

Figure 9 Load profiles of the subway network in a typical week

CONCLUSION

This paper has shown several real-world applications of how smart card fare validation data are used within a transit authority setting. System intelligence from smart card provides different units within the organization and across partner organizations in the region access to up-to-date and previously unavailable information. This in turn allows them to make more informed decision on policies and plans. The future of public transit relies greatly on adequate data and intelligence. The next steps would be to expand the number of applications, to put in place a structure that facilitates systematic and corporate use of the data and to increase cooperation with partner organizations.

REFERENCES

Chu, K.K.A. and R. Chapleau. Enriching Archived Smart Card Transaction Data for Transit Demand Modeling. In Transportation Research Record: Journal of the Transportation Research Board, No. 2063, TRB, National Research Council, Washington, D.C., 2008, 63-72.

Chu, K.K.A., R. Chapleau and M. Trépanier. Driver-Assisted Bus Interview: Passive Transit Travel Survey with Smart Card Automatic Fare Collection System and Applications. In Transportation Research Record: Journal of the Transportation Research Board, No. 2105, TRB, National Research Council, Washington, D.C., 2009, 1-10.

Chu, K.K. Leveraging data from a smart card automatic fare collection system for public transit planning. Ph.D. Thesis, École Polytechnique de Montréal (Université de Montréal), Montréal, Q.C., Canada, 2010.

Chu, K.K.A. and R. Chapleau. Augmenting transit trip characterization and travel behavior comprehension: multi-day location-stamped smart card transactions. Transportation Research Record: Journal of the Transportation Research Board, No. 2183, 2010, 29-40.

Chu, K.K.A. and D. Bergeron. Exploiting data of a multi-modal and multi-operator smart card automatic fare collection system from the perspective of a regional transit authority. Paper presented at the 12th WCTR conference, Lisbon, Portugal, 2010.

Chu, K.K.A. and R. Chapleau. Smart card validation data as a multi-day transit panel survey to investigate individual and aggregate variation in Travel Behaviour. In J. Zmud, M. Lee-Gosselin, M. Munizaga and J. A. Carrasco (Eds.), Transport survey methods: Best practice for decision making (649-671), 2013. Bingley: Emerald.

Chu, K.K.A., T. Spurr and R. Chapleau. Travel-demand characterization for a large urban transit infrastructure using smart-card transaction data. Paper presented at the 13th WCTR conference, Rio de Janeiro, Brazil, 2013.

Chu, K.K.A., Characterizing long-term transit service consumption and fare usage patterns with disaggregate smart card data. Paper presented at the 93rd meeting of the Transportation Research Board, Washington, D.C., 2014.

Lee, S. and M. Hickman. Travel pattern analysis using smart card data of regular users. Paper presented at the 90th meeting of the Transportation Research Board, Washington, D.C., 2011.

Ma, X., Y.-J. Wu, Y. Wang, F. Chen and J. Liu. Mining smart card data for transit riders' travel patterns. Paper presented at the 92nd meeting of the Transportation Research Board, Washington, D.C., 2013.

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

Park, J.Y., D.J. Kim and Y. Lim. Use of the smart card data to define public transit in Seoul, South Korea. Transportation Research Record: Journal of the Transportation Research Board, No. 2063, 2008, 3-9.

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

Sammer, G. Identifying and reconciling the data needs of public transit planning, marketing and performance measurement. In P. Bonnel, M. Lee-Gosselin, J. Zmud & J.-L.Madre (Eds.), Transport survey methods: Keeping up with a changing world (321-348), 2009. Bingley: Emerald.

Utsunomiya, M., J. Attanucci and N. Wilson. Potential uses of transit smart card registration and transaction data to improve transit planning. Transportation Research Record: Journal of the Transportation Research Board, No. 1971, 2006, 119-126.

DISCLAIMER

The views and opinions expressed in this paper are solely of the author and do not necessarily reflect the official policy or position of the AMT. Examples of analysis presented in this paper are for illustrative purpose only and are not intended for official use.