Mining Smart Card Data from an Urban Transit Network

Bruno Agard
École Polytechnique de Montréal, Canada

Catherine Morency
École Polytechnique de Montréal, Canada

Martin Trépanier
École Polytechnique de Montréal, Canada

INTRODUCTION

In large urban areas, smooth running public transit networks are key to viable development. Currently, economic and environmental issues are fueling the need for these networks to adequately serve travel demand, thereby increasing their competitiveness and their market share. Better balance between transit supply and demand will also help reduce and control operating costs.

The fact is, however, that transit operators are finding it extremely difficult to adjust the service to meet the demand, because this demand changes continuously with the time or day of travel (period of the day, day of the week, season or holiday) and other factors like weather and service breakdown. In order to enhance their service, operators need to better understand the travel demand (customer behaviors and the variability of the demand in space and time). This can be achieved only by continuously monitoring the day-to-day activities of users throughout the transit network.

Some large cities around the world take advantage of smart card capabilities to manage their transit networks by using Smart Card Automated Fare Collection Systems (SCAFCS). An SCAFCS gives travelers greater flexibility, since a single card may be used by one user at various times and on different parts of the transit network, and may support various fare possibilities (by travel, line, zone, period, etc.). For transit operators, these systems not only validate and collect fares, but also represent a rich source of continuous data regarding the use of their network. Actually, this continuous dataset (developed for fare collection) has the potential to provide new knowledge about transit use. Following the application of various pretreatments which make it possible to extract real-time activity, data mining techniques can reveal interesting patterns. These techniques are aimed at precisely describing customer behavior, identifying sets of customers with similar behaviors, and measuring the spatial and temporal variability of transit use. Patterns are extracted and analyzed to document various issues, such as identifying transit use cycles or homogeneous days and weeks of travel for various periods of the year. This information is required for a better understanding and modeling of customer behavior, and consequently better adjustment of the service to the demand. These adjustments may, for instance, lead to the restructuring of the transit network, to the adaptation of route scheduling or to the definition of new and different subscription options (fares).

Below, results from various experiments conducted with a real dataset are provided. They show the potential of data mining to provide useful and novel information about user behavior on a transit network. The data processed in the study are extracted from a system operating in a Canadian city (Gatineau, Quebec).

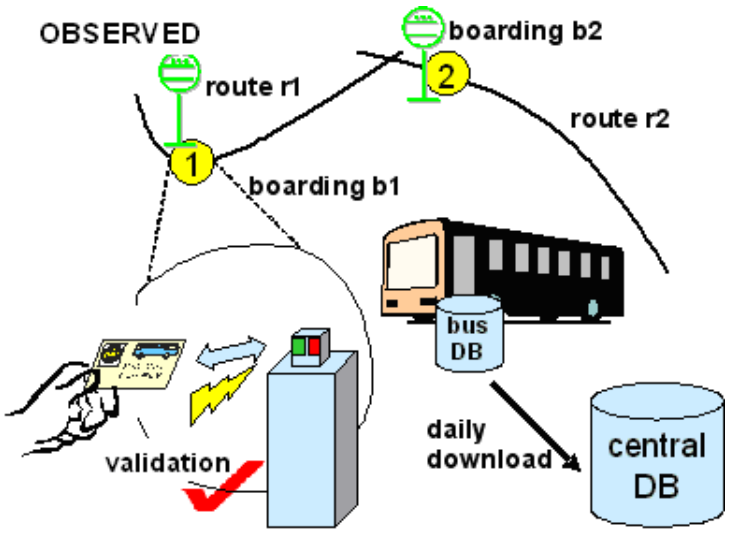
BACKGROUND

Smart Card in Public Transport

Generally, SCAFCS are composed of cards, onboard readers and a centralized information system (see Figure 1).

A smart card is simply an RFID device implanted in a transport card, and is similar to a typical credit card. Smart cards offer various advantages over traditional paper systems:

Figure 1. Smart card automated fare collection system (SCAFCS)



- Validation of the user transaction (boarding/transfer) is instantaneous and does not require any interaction with the driver.
- Complex fare systems (multiple zones, rates), common in large metropolitan areas where networks are integrated, can easily be managed.
- Elementary security procedures can be taken.
- Data are continuously collected, resulting in datasets much larger than those usually available to measure customer traffic; hence, a larger user proportion is observed.

These systems are being used increasingly frequently, since the software and hardware tools necessary to support their implementation are now stable and accessible, with a good quality of data output (Chira-Chavala and Coifman 1996, Meadowcroft 2005). Smart cards in transportation have mainly been implemented to simplify fare collection, but could also be used to advantage to monitor the service itself. In most cases, the adoption of a smart card system by transit authorities is related to their level of funding and the level of sophistication of their other technologies (Iseki et al. 2007, Yoh et al. 2006). Cheung (2006) demonstrated the overall benefits of smart card systems in The Netherlands, but it was Bagchi and White (2004, 2005) who were the first to substantiate this potential for transit planning. Using three case studies (British networks), they illustrate the ability of smart card data to estimate turnover rates, trip rates per card and the impacts of the use of smart cards on the number of linked trips. They also discuss the complementary nature of smart card data and other data collection methods, arguing that smart cards should not replace those methods. Utsunomiya et al. (2006) presented a study based on Chicago Transit Authority data for a one-week period involving about 500,000 boarding transactions. They reported difficulties associated with the data, especially missing transactions and incorrect bus routes.

Travel Behavior Variability

In practice, many transit planners use statistics from synthetic models, onboard travel surveys or regional travel surveys to describe how their networks are used. Average statistics are constructed, describing typical customer behaviors during a typical weekday. The underlying hypothesis that all weekdays are similar is