LEARNING MOBILITY USER CHOICE AND DEMAND MODELS FROM PUBLIC TRANSPORT FARE COLLECTION DATA

Frederic Roulland, Luis Ulloa, Arturo Mondragon, Michael Niemaz, Guillaume Bouchard, Victor Ciriza
Xerox Research Centre Europe
6 Chemin de Maupertuis
38240 Meylan, France
+33 4 76 61 50 50, firstname.lastname@xrce.xerox.com

ABSTRACT
In this paper we present a new approach for public transit simulation which is based in machine learning techniques in order to model user choices and demand. We believe this will enable a new generation of much finer grained simulations for planning public transit infrastructure development making use of the richness of the information contained in the mass of data collected by modern transportation systems. We describe how we have implemented the system using public transport fare collection data, what is our current results and our next steps for the maturation of this technology.

INTRODUCTION
Improving the transportation of urban areas is a constant challenge for any transportation authority. However development of infrastructure is costly and heavy work for a city and it therefore requires a very careful study of the mobility needs and simulation of different alternatives before taking such decision.

Today the planning of a new transportation infrastructure involves simulation of several options based on an understanding of the city which is mostly based on a custom study. This study generally consists on a period of manual collection of data and surveys which are then aggregated into a standard model of the city by an expert analyst. This approach is expensive and therefore it often has the disadvantages to rely only on partial data (only few thousands of people surveyed and often with poor time coverage) and/or non up-to-date data (e.g. global census of 5 years ago). Furthermore mapping this data into a model is such an enormous work that data are often aggregated to a high level of granularity which allows only a macro simulation of the city mobility dynamic. Although tailored, standard models are taken and assumed to hold for various locations such as Stockholm and Mexico City. Therefore we believe that the cost/precision ratio of simulations built from these custom studies is not optimal.

On the other hand, the presence of more and more intelligent transportations systems on the field allows the collection of millions of transactions which constitute a very detailed source of information. This is for example, public transportation ticketing data that are collected from automatic fare collection systems each time a passenger checks-in (and sometimes checks-out) before boarding (alighting) a vehicle. So far it has been very limitedly used for understanding mobility patterns of a city (see (1, 2)). We propose a mechanism that will use this data in order to learn automatically city models that can then be used for simulation and
planning of new infrastructure. We therefore hope to increase significantly the planning cost/precision ratio through the integration of a much more massive, more up to date set of information, in an automated way.

**SYSTEM OVERVIEW**

At the core of our system, there is a learning module that is able to model the city mobility without any human intervention just by using the data previously collected by the fare collection system. This module relies however heavily in two sub-components:

The passenger journey reconstruction module’s role is to reconstruct all the journeys of each individual users of the public transit service from the logs of tickets/card validations. It relies on the work presented in (3) which is based on the assumptions that the number of travellers that follow a symmetrical pattern and use the same ticket/card are characteristic of the demand. This module generates in essence what will constitute together with the schedules and lines description of the current service the training set for the learning module.

A micro-simulation module constitutes the other main sub-component. This module is able to generate a set of traces that represents all the simulated journeys of each individual public transit passenger for a given period of time. It uses some automata simulating each individual traveller and each public transit vehicle within the city network and their interaction. This micro-simulator takes as a configuration input the definition of schedules and lines of the public transit service. This is used to populate all the vehicles and to simulate their behaviours. This is used as an initial point to learn the probability of schedule adherence. It also takes as an input the city mobility model that is being learned by the learning module. This model is composed basically of two parts. One part models the demand to be learned, i.e. the origins and destinations at any time of the travellers. The micro-simulator uses this part to populate all the travellers and their journeys in the simulation. The second part models the travellers’ behaviour to be learned, i.e. what will drive the choice of a particular path given the origin, the destination, the context and the profile of a traveller. This user choice model relies heavily on a trip planner which has been designed specifically for computing very quickly a large set of alternatives for a given origin and destination at a given time. It is called during a simulation each time a traveller automata needs to simulate the choice of a traveller starting a journey or having to reroute his journey because of an unexpected change.

So far very few approaches have attempted to use a micro-simulation in order to simulate an entire public transit network (4). We believe this is mainly due to two constraints. The main one is the labour cost that would be associated to collect all the data and to build manually the mobility model of the city. The second one is associated to the computational complexity of running the simulation. Whilst our approach of automated learning of the model solves the first issue we have as well implemented custom algorithms to allow micro-simulation and trip planning in a distributed environment.

**CURRENT RESULTS AND NEXT STEPS**

The system has been fully implemented and is currently tested with some fare collection data provided by the ATLAS system [5] which is deployed in the city of Nancy, France. This city operates a network with daily journeys in the order of 100 000. The results of the simulation can be analysed using the data mining tool developed on top of ATLAS. The results are
indeed exactly of the same format than the validations logs collected by ATLAS. For illustration purposes we can show for example in Figure 1 using this tool what is the aggregated load of the public transit network over a day of activity in the city of Nancy and what is, in contrast, the result of a simulation where we tested what would happen if the tram line acting as backbone in the network would be removed. We can see from the two different visualisations how the removal of the line is partly reducing the overall use of the public transit but also how in some places some parallel routes are experiencing major load increases.

![Heatmaps for Nancy public transit network daily load (left) and simulation of load without tram line (right)](image)

Figure 1: Heatmaps for Nancy public transit network daily load (left) and simulation of load without tram line (right)

Our work is still preliminary and it only demonstrates the technical feasibility of such approach. We are currently working with some historical data of public transportation networks where some significant changes were made. We are using this data in order to evaluate the accuracy of our simulations and to test several models in order to find one which combines quick convergence properties during the learning phase and good robustness for accurate prediction in the simulation of various changes in the public transit service. Along this path we believe that one of the natural next steps in the evolution of this learning process will be to be able to take into account additional data than the public transit tickets validations for capturing a larger view of the mobility patterns of a city.

REFERENCES


