Learning-based Departure Time and Path Choice Modelling for Transit Assignment under Information Provision: A Theoretical Framework
Mohamed Wahba* and Amer Shalaby

Keywords: Passenger Assignment, Learning, Markovian Decision Process

This paper presents the theoretical development of a departure time and transit path choice model based on the Markovian Decision Process. This model is the core of MILATRAS - Microsimulation Learning-based Approach to TRansit ASsigmnet. Wahba and Shalaby (2009) report on a full-scale implementation of MILATRAS, including the proposed departure time and path choice model of this paper, using the Toronto Transit Commission (TTC) system as a case study.

Recently, traffic assignment procedures implemented learning-based models which have been shown to result in different and more realistic assignments relative to conventional methods (e.g. Ettema et al., 2005). Similar findings have been presented for modelling traffic assignment models with information provision using multi-agent representation (e.g. Rossetti, 2005). As such, the next generation of dynamic transit assignment algorithms should consider adopting learning-based and multi-agent microsimulation concepts for consistency with emerging traffic assignment models and for integration with state-of-the-art activity-based modelling frameworks.

Passengers, while travelling, move to different locations in the transit network at different points in time (e.g. at stop, on board), representing a stochastic process. This stochastic process is partly dependent on the transit service performance and partly controlled by the transit rider. This can be analyzed as a Markovian Decision Process (MDP). In an MDP, actions are rewarded and hence passengers’ optimal policies can be estimated. For example, for an origin stop choice, there is an immediate cost expressed as the value of the access cost (time and/or money). Also, there is a future value of a specific stop choice expressed in the expected travel cost of the subsequent available route and transfer connection choices. Similarly, for a route choice, there is an immediate cost expressed as the value of waiting time, while future cost is related to the value of possible transfer connections and the probability of arriving at the desired arrival time.

The Markov Chain representing the transit path choice problem is shown to be ergodic. In the transportation context, this means that each possible combination of path choices can be tried by iterative dynamics (i.e. all path options can be tried through making the same trip repeatedly over days).

Travellers are assumed to be goal-directed intelligent agents. A traveller has a goal of minimizing the travel cost (e.g. time, money, inconveniences). In order to achieve their goals, travellers follow policies to optimize the return of their trip. Assuming passengers are rational, it is logical to expect passengers to follow their optimal policy and to optimize their trip return (or cost in this regard). The effect of such an optimal policy is observed through either disaggregate

---

1 Postdoctoral Associate, Department of Civil and Environmental Engineering, Massachusetts Institute of Technology, RM 1-153 – 77 Massachusetts Avenue, Cambridge MA 02139, mwahba@mit.edu
* Corresponding author
2 Associate Professor, Department of Civil Engineering, University of Toronto, 35 St. George Street, Toronto, Ontario, M5S 1A4, amer@ecf.utoronto.ca
individual choices, and/or aggregate route loads. If the underlying Markov process is ergodic, then there exists a unique optimal policy; if followed, passengers will optimize their return, associated with is a steady-state transition probability matrix. Passengers devise their optimal policies based on a value function for the state-action evaluation. While the optimal policy is observed (or its effect through route loads), the only unknown (to the modeller) is the value function. By reconstructing the transit path choice problem as a Markovian Decision Process, the value function (or its parameters), used by individual passengers, can be estimated/calibrated. This is similar to the process of Inverse Reinforcement Learning (IRL).

The underlying hypothesis is that individual passengers are expected to adjust their behaviour (i.e. trip choices) according to their experience with the transit system performance. Individual passengers base their daily travel decisions on the accumulated experience gathered from repetitively travelling through the transit network on consecutive days. Travellers’ behaviour, therefore, is modelled as a dynamic process of repetitively making decisions and updating perceptions, according to a learning process. This decision-making process is based on a mental model of the transit network conditions. The learning and decision-making processes of passengers are assumed in the proposed model to follow Reinforcement Learning (RL) principles for experience updating and choice techniques.

The proposed learning-based choice model considers the departure time choice, the stop choice and the run (or sequence of runs) choice. The model is appropriate for modelling information provision since it distinguishes between individual’s experience with the service performance and information provided about the system dynamic characteristics. Future work includes the integration of a mode choice component into the overall modelling framework, using the learning-based approach. Research efforts will also be directed to the modelling of passengers’ travel choices in a multimodal network, incorporating the access and egress mode choices. Finally, the growing trend in using smart cards (e.g. new GTA smart card fare system “Presto”, Chicago Card Plus, and the Charlie Card in Boston) provides a rich source of data on transit departure time and path choices in particular and multi-modal trips in general. Such data sources will present great opportunities for calibrating and validating dynamic transit path choice models.

REFERENCES

