Journal of Computational Science 10 (2015) 338–350

Contents lists available at ScienceDirect

Journal of Computational Science

journal homepage: www.elsevier.com/locate/jocs

A data-driven agent-based model of congestion and scaling dynamics of rapid transit systems☆

Nasri Bin Othman, Erika Fille Legara, Vicknesh Selvam, Christopher Monterola*

Complex Systems Programme, Institute of High Performance Computing, Agency for Science, Technology and Research, 1 Fusionpolis Way, #16-16, Connexis North, Singapore 138632, Singapore

ARTICLE INFO

Article history:
Received 13 October 2014
Received in revised form 19 March 2015
Accepted 20 March 2015
Available online 24 March 2015

Keywords:
Agent-based modelling
Data-driven simulations
Urban complexity
Transportation dynamics
Smart card data

ABSTRACT

Investigating congestion in train rapid transit systems (RTS) in today's urban cities is a challenge compounded by limited data availability and difficulties in model validation. Here, we integrate information from travel smart card data, a mathematical model of route choice, and a full-scale agent-based model of the Singapore RTS to provide a more comprehensive understanding of the congestion dynamics than can be obtained through analytical modelling alone. Our model is empirically validated, and allows for close inspection of congestion and scaling dynamics. By adjusting our model, we can estimate the effective capacity of the RTS trains as well as replicate the penultimate station effect, where commuters travel backwards to the preceding station to catch a seat, sacrificing time for comfort. Using current data, the crowdedness in all 121 stations appears to be distributed log-normally. We find that increasing the current population (2 million) beyond a factor of approximately 10% leads to an exponential deterioration in service quality. We also show that incentivizing commuters to avoid the most congested hours can bring modest improvements to the service quality. Finally, our model can be used to generate simulated data for statistical analysis when such data are not empirically available, as is often the case.

1. Introduction

To tackle rising population density in urban cities, transportation planners often construct train rapid transit systems (RTS) as a first step. Yet continued population growth forces the RTS to evolve towards increased complexity with more train lines added to satisfy demand. With the increased complexity, planners are confronted with the difficulty of predicting commuter ridership, route choices, and also the various outcomes of the system during disruptions. Moreover, increased station and train crowdedness in RTS lead to congestion, commuter discomfort, trip delays, and lowered overall service quality standards. It is therefore imperative that modern transportation models be capable of investigating not just the issues of efficient, robust and scalable transportation, but also of commuter comfort and satisfaction.

The introduction of smart card ticketing in RTS has serendipitously enabled large-scale data analytics into commuter travel behaviour [1,17]. Analytical and regression models have been developed to estimate commuters' spatio-temporal density [20], identification of boarded trains [10], travel patterns [4], and transit use variability [14]. Yet, it is noted that the information captured by smart cards has limitations [17]; for example, most datasets do not contain routing information as they capture information only at the entry and egress points of journeys.

In contrast to analytical and regression models, agent-based models (ABM) strive to model each individual agent in a manner most natural to the system at hand [3]. Essentially, an agent is autonomous and formulates decisions and interacts with other agents directly. By directly replicating the mechanics of the system, an ABM permits the observation of emergent phenomena that arise from the interactions of the agents with each other [3] – provided the mechanics are correctly characterized and the model is well-calibrated.

ABM has seen recent success in modelling large-scale transportation [7,15,21]. However, there are not many studies which incorporate smart card data to drive RTS demand for better calibration. In our previous work [11], we had leveraged upon anonymized travel smart card transactional data to synthesize travel demand for a smaller-scale agent-based model of the Singapore transit system involving only one of the operational train lines, and achieved a very close match between the simulated and empirical travel duration distributions. In that work, we also investigated the dynamics of the smaller-scale system with regard to population growth.

☆ This article is an extension of [16].

* Corresponding author. Tel.: +65 64191554; fax: +65 64631452.
E-mail addresses: othmannb@ihpc.a-star.edu.sg (N.B. Othman), legaraeft@ihpc.a-star.edu.sg (E.F. Legara), selvamv@ihpc.a-star.edu.sg (V. Selvam), monterolac@ihpc.a-star.edu.sg (C. Monterola).

http://dx.doi.org/10.1016/j.jocs.2015.03.006
1877-7503/

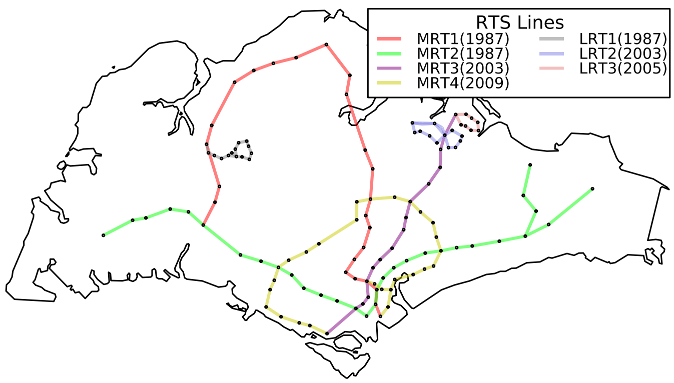

Fig. 1. Singapore rail transit system. The four MRT lines and three LRT lines spanning 121 stations are shown. For each line, the year of first operation is shown in parentheses in the legend.

Here, we extend our previous work [11] by: (1) expanding the model to cover all seven operational lines; (2) adding a route-choice mechanism inferred statistically from empirical travel duration distributions [13]; (3) incorporating station-specific walk-times; (4) investigating dynamics that were not directly measurable in our dataset, such as station crowdedness; (5) estimating the effective train capacity; (6) modelling the penultimate station effect; and (7) running further population scaling scenarios. We validate our model by ensuring the travel duration distributions generated from our simulations are well-calibrated to the empirical dataset. This would lend strength to any inferences derived from our scenarios. Apart from these goals, ultimately, we strive to construct a simulation platform that can be used to evaluate the efficacy of proposed strategies in tackling current and future urban transportation issues.

Our experiments in this work are focused on the Singapore rail transit system, which began operations in 1987 and is now one of the busiest RTS in the world. Despite the focus, our approach can be applied to other rail transit systems in the world, as we do not utilize any Singapore-specific mechanics or assumptions in our model.

2. Data

Our model is dependent on data for the following purposes: (1) to construct the transit infrastructure, (2) to instantiate the commuter agents corresponding to the actual travel demand, (3) to calibrate the travel time components of the network, and (4) to accurately model the commuters' decision making (e.g., route choice).

We model the Singapore rail transit system in our experiments. The Singapore RTS comprises two train systems: the Mass Rapid Transit (MRT) system consisting of four standard-gauge lines spanning 90 stations, and the Light Rapid Transit (LRT) system consisting of three regional tram lines spanning 31 additional stations, as shown in Fig. 1. To serve the commuting demands of the city, the service frequency of the trains are kept from 2 to 4 min during peak hours, and from 4 to 8 min during off-peak hours. The tapping gates at the stations are typically located within 50 m of the station platforms. This distance introduces a walking time component in the commuter travel duration. Interchange stations which connect multiple lines contain internal walkways where commuters can transfer from one platform to another platform directly without having to first leave the station.

To construct the train network in our model, we consulted publicly available resources, including the LTA website.[1] This yields the set of all train stations, their connectivity, and estimated travel time between two adjacent stations. Information regarding the first and last trains at each station are also publicly available, and is used to estimate the train dispatch schedule.

To estimate the walking times along the stations in order to account for commuter locomotion, we conducted several physical tours of the stations. The measurements made are coarse and not empirically verified; however, they are sufficient since walking is usually the smallest component of travel – typically less than 2 min.

Our main data source for the commuter travel demand is the anonymized travel smart card dataset for public transport users in Singapore, obtained from the Land Transport Agency (LTA) of Singapore. This amounts to over 14 million train journey records for 2 million anonymized card IDs taken across a full week. Note that the population of Singapore is approximately 5.3 million in this period. Thus, over one-third of the Singapore population ride the public trains. In our data, a trip begins with a tap in of the smart card at the origin station, and terminates with a tap out at the destination station. Here, we use the following fields for each record: *origin (tap-in station)*, *destination (tap-out station)*, *tap-in time*, and *trip duration*. From the *origin*, *destination*, and *tap-in time* fields, we can reconstruct the travel demand for any given origin–destination (O–D) pair and time. The *trip duration* field is used for validating the simulation.

3. Computational model

Our approach to modelling the RTS comprises two aspects: (1) the modelling of the trains as they traverse the rail network, and (2) the modelling of commuters as they travel from their origins to their destinations. The first aspect, the modelling of trains, is straightforward as we do not fully model the physical mechanisms of the trains, and it is only sufficient that our simulated train arrivals can fit the publicly available train schedules (i.e., first-train timings and train arrival frequencies for each station), and that our train capacities are estimated correctly. Modelling commuters however, is non-trivial, as we seek to capture the totality of the experiences of every commuter, from the moment they tap in at the origin, to the moment they tap out at the destination; and our commuter population is heterogeneous in physical characteristics and decision preferences. To serve both aspects, we have chosen to model both the trains and commuters as agents in a time-based simulation spanning a day of service. As time progresses, train agents are launched on the station platforms and traverse the rail network according to their simulated schedules. Concurrently, commuter agents are created at origin stations and board the passing trains to their destination stations where they will leave the system. The life-cycles of these two types of agents are captured in the state diagrams in Fig. 2.

3.1. Train modelling

The rail network in Singapore – comprising 121 stations, 412 directed edges connecting adjacent stations, and 7 train lines – can be depicted using a directed graph (V, A), consisting of vertices representing the stations, V, and arcs connecting adjacent stations together, A. The cost of each arc represents the time it takes for a train to travel from one station to an adjacent station. Although connectivity between adjacent stations is symmetric, the travel time is not; therefore the graph is directed. The set of arcs can be further categorized into several disjoint sets representing the 7 operational lines in the rail network. The physical attributes of a train depends on which line it is on, as some lines utilize smaller capacity trains than the others (1920 being the typical train capacity).

Since a train will wait at a station while commuters board and alight it, we also need to model the individual platforms for each station. For every train which passes by a station S, it has an arc that

[1] http://www.publictransport.sg/content/publictransport/en/homepage/trainmap.html (last accessed January 2014).