Land Use Policy 57 (2016) 356–367

Contents lists available at ScienceDirect

Land Use Policy

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

Impacts of land use and amenities on public transport use, urban planning and design

Nan Hu[a], Erika Fille Legara[a], Kee Khoon Lee[a], Gih Guang Hung[a,b], Christopher Monterola[a,c,*]

[a] Institute of High Performance Computing, 1 Fusionopolis Way, #16-16 Connexis North, Singapore 138632, Singapore
[b] Rolls-Royce Singapore Pte. Ltd., Advanced Technology Centre, 6 Seletar Aerospace Rise, Singapore 797575, Singapore
[c] Complexity Institute, Nanyang Technological University, 50 Nanyang Avenue, Singapore 639798, Singapore

ARTICLE INFO

Article history:
Received 8 September 2015
Received in revised form 24 May 2016
Accepted 1 June 2016

Keywords:
Land use
Amenities
Public transit
Urban planning

ABSTRACT

Various land-use configurations are known to have wide-ranging effects on the dynamics of and within other city components including the transportation system. In this work, we particularly focus on the complex relationship between land-use and transport offering an innovative approach to the problem by using land-use features at two differing levels of granularity (the more general *land-use sector types* and the more granular *amenity structures*) to evaluate their impact on public transit ridership in both time and space. To quantify the interdependencies, we explored three machine learning models and demonstrate that the decision tree model performs best in terms of overall performance—good predictive accuracy, generality, computational efficiency, and "interpretability". Results also reveal that amenity-related features are better predictors than the more general ones, which suggests that high-resolution geo-information can provide more insights into the dependence of transit ridership on land-use. We then demonstrate how the developed framework can be applied to urban planning for transit-oriented development by exploring practicable scenarios based on Singapore's urban plan toward 2030, which includes the development of "regional centers" (*RCs*) across the city-state. Results show an initial increase in transit ridership as the amount of amenities is increased. This trend, on the other hand, eventually reverses (particularly during peak hours) with continued strategic increase in amenities; a tipping point at 55% increase is identified where ridership begins to decline and at 110%, the predicted ridership begins to fall below current levels. Our *in-silico* experiments support one of the medium-term land-use transport goals of stakeholders—to alleviate future strains in the transportation system of Singapore through the development of *RCs*. The model put forward can serve as a good foundation in building decision-support tools that can assist planners in better strategizing and planning land-use configurations, in particular the amenity resource distribution, to influence and shape public transportation demand.

1. Introduction

With rapid urbanization, issues on urban sustainability and resilience have become more and more challenging to synergize with multiple, and sometimes opposing, objectives (Filion and McSpurren, 2007). As a consequence, it is essential to probe multiple urban indicators in a more systematic and holistic manner to capture various urban-related phenomena such as transport ridership and road traffic flow—those that are known to be influenced by the interactions of the different components of an urban system (e.g., land-use, transport, population, etc.). Understanding the interplay of these factors is vital to accurately evaluate existing conditions "on-ground" and to effectively determine how to plan and design urban systems including the identification of which public infrastructures, services, and resources need to be built and deployed (Batty, 2007).

Quantifying public transit ridership involves complex processes as it is affected by various factors: socio-economic (Wang, 2011; Greer and van Campen, 2011), geographic and spatial (land-use related) (Litman, 2007; Putman, 2013; Beimborn et al., 1999; Choi et al., 2012; Engelen, 1988), financial (public) (Neil et al., 2006; Bettencourt, 2013; Lee et al., 2015), and other more qualitative factors such as comfort and convenience (Chen et al., 2011; Litman, 2008). Most of these factors, however, can be quite laborious to

* Corresponding author.
E-mail addresses: hun@ihpc.a-star.edu.sg (N. Hu), legaraeft@ihpc.a-star.edu.sg (E.F. Legara), leekk@ihpc.a-star.edu.sg (K.K. Lee), terence.hung@rolls-royce.com (G.G. Hung), monterolac@ihpc.a-star.edu.sg (C. Monterola).

http://dx.doi.org/10.1016/j.landusepol.2016.06.004
0264-8377/

measure, more so to control vis-à-vis urban planning. Hence, in this work, we particularly zoom into the more tangible land-use features to quantify ridership. Although these physical features are less challenging to appraise, the relationship of land-use and transport is not any less complex. On one hand, land-use design and characteristics have been used to build various travel demand models (TDM) (e.g., the Four Step TDM (Manheim, 1979; Florian et al., 1988)) as land-use features impose specific spatial constraints for most, if not all, activities. On the other, the demand for transport within an area influences the price of land and the distribution of amenities within a locality, which then affects the future evolution of land-use design (Beimborn et al., 1999; Still et al., 1999; Decraene et al., 2013).

Traditionally, at the planning level, studies have been more macroscopic in nature—without detailed spatiotemporal analysis of the ridership at individual localities (Choi et al., 2012; Chakraborty and Mishra, 2013). Here, we offer a fresh approach by zooming in on certain planner controllable land-use features at higher resolutions and by analyzing their respective spatiotemporal impacts on ridership from both the aggregated and more localized levels. We believe that from an urban planning perspective, it is more suitable to focus on controllable features such as total population, transportation infrastructure, and land-use than factors at lower operational levels (e.g., people's perception on the use of alternative transport mode (Taylor and Fink, 2002)). Our approach makes use of travel data collected from an automated fare collection system, which is used together with certain geo-datasets such as amenity types and distributions obtained from OpenStreetMap (OSM, 2015)—an open source map database. Using open source data, the methods developed and proposed can be easily extended and applied to other countries and/or cities.

We then introduce three multivariate analytical models to quantify the relationship between a set of land-use features and public transit ridership. In multivariate analysis, collinearity is a big challenge (Gomez-Ibanez, 1996; Crane, 2000) as variables can be inter-correlated (e.g., residential and industrial estate can be separated naturally as the urban system evolves (Decraene et al., 2013)); thus, we implement different multivariate analytical models in the estimation of ridership. Our goal is to not only find the most suitable model that would result to the highest accuracy rates, but to also identify the relative importance of the different variables and their critical values in the forecasting process. The machine learning methods, therefore, are evaluated based on four criteria: (1) predictive accuracy; (2) generality in handling different constraints and assumptions of unknown variable values; (3) computational cost/complexity; and (4) "interpretability". To demonstrate the use of our best model, we apply it to tangible urban development scenarios, including the development of Regional Centers (*RC*s), to quantify the effects of the plans to the ridership.

The rest of the paper is organized as follows. In the next section, we discuss the different multivariate analytical models used to quantify the interrelation between various land-use features and public transit ridership. Section 3 then evaluates the predictive accuracy, generality, and ease of interpretation of these methods. Scenario studies on hypothetical plans that include amenity resource increments around *RC*s are implemented and discussed in Section 4. Finally, in Section 5, we conclude the paper and provide helpful insights into our work's usability to urban planners and other stakeholders.

2. Multivariate models and evaluations

First, we define a *locality* as a surrounding area centered at a public transport station (bus or train). The ridership within a locality is the number of individuals going in and out of the station as recorded in the anonymized electronic ticketing cards. The transport demand at a given station (or stop) over time within a day gives us an idea on the role of the transport point in the entire system. We then infer that the utilization of land-use entities surrounding a station is linked to its ridership as what we have described in our previous work on city characterization based on different land-use category data (Decraene et al., 2013). We then implement three (3) machine learning models to reinforce this connection between land-use features and ridership.

2.1. Urban data in Singapore

The primary data utilized in this study are as follows:

- **Singapore URA Master Plan 2008 (MP2008)**. Land-use categories were extracted from the government's master plan on land-use allocation (Fig. 1(a)). In the original MP2008, the land-use categories include more detailed categories, where we aggregated similar categories (such as different types of business sector) into five broader sectors[1]: business, industrial, residential, water, and others. The original resolution of the image map is 9.7×9.7 sqms per pixel; the map was then merged and scaled down to 31.25×31.25 sqms per pixel (approximately 32 pixels per 1 km) to make it consistent with other complimentary datasets.
- **Greenery**. The greeneries dataset was extracted from Landsat 7 satellite multi-spectral imagery dated 2002 with less than 5% cloud coverage. The original resolution is 31.25×31.25 sqms per pixel. The density of greeneries is not discriminated in this dataset. This dataset complements the MP2008 such that the intersections between the greenery area in this map and the "others" land-use category is considered as a "greenery" (land-use type). Consequently, we now have six sectors of land-use: *residential*, *business*, *industrial*, *greenery*, *water*, and *others*.
- **Amenities**. The amenities of a locality are used to add more granularity to its land-use features, in addition to the land-use category data, see Fig. 1(d). Amenity information was retrieved from Open Street Map (OSM)—an open source geo-data platform that provides a free map of the world. At least for Singapore, we have verified that the geoinformation in OSM is accurate. Based on the usage of the amenities and their tag information in OSM, the amenities are grouped into eight categories: *sustenance*, *education*, *transportation*, *healthcare*, *entertainment*, *finance*, *commerce* and *others*. The same scale of 31.25×31.25 sqms per pixel was applied in calculating the amenity densities.
- **Transport data**. The smart fare card dataset used in this work was collected from a centralized automated fare collection (AFC) system and was provided by the Singapore Land Transport Authority (LTA). In recent years, "contactless" smart fare cards and the travel information they generate have particularly helped advance research in the field of human mobility, transport, and urban planning, among others, allowing researchers to probe various aspects of the dynamic spatiotemporal patterns that commuters generate within a given territory in a less invasive manner (Legara and Monterola, 2015; Hasan et al., 2013; Pelletier et al., 2011; Bagchi and White, 2005). In this work, for example, we utilize travel data to inform and train our supervised machine learning models with ridership demand as model output. For the purpose of our research, we only considered the following travel information generated from the AFC system: total number of commuters

[1] More details on the land sector aggregation process are reported in our previous work (Decraene et al., 2013).