PATREC

Planning and Transport Research Centre (PATREC)

EXECUTIVE SUMMARY

Travel Behaviour Patterns – Macro Analysis

Project No	Project 4.1
Authors	Simon Moncrieff, Tristan Reed, S Zaung Nau, Sharon Biermann
	Damien Martin (DOP), Wes Soet (MRWA), Graham Jacoby (MRWA), Charlotte Hayes (PTA), Sue Hellyer (DOT), Rosie Selby (DOT), Renlong Han (DOT), Laura Cook (Treasury)
Date	November 2017

The University of Western Australia · ECU Edith Cowan University · Curtin University · Main Roads Western Australia · Western Australian Planning Commission · Department of Transport

Executive Summary

Project 4 Background

Project 4 is an evidence based, data driven project centred around the analysis of SmartRider data. The project is divided into two complementary projects, one focusing on macro analysis (4.1), investigating large scale patterns in the data, and the other on micro analysis (4.2), studying the patterns of individuals.

Project 4.1 Purpose: This project aims to bridge the gap between data and modelling output. A generalised data mining framework for analysing travel behaviour patterns was developed. The framework included the ability to incorporate explanatory variables beyond transport, such as socio-economic factors, and support the generation of indicators along with the automation of data modelling.

Project 4.2 Purpose: This project takes a new, data-driven, approach for learning customer profiles from their actual travel patterns. For example, what are the different types of customers? What are the driving factors for hub utilisation? To answer these types of questions a system is being developed for querying, analysis and data mining, to support the overall knowledge discovery process centred on customers, hubs, and journeys. The project will deliver analysis methods for discovering patterns and software tools to facilitate that analysis and for visualisation of the results for policy makers.

Research Question(s)

- What are the spatial-temporal transport usage patterns as evident from SmartRider data in relation to socio-economic factors?
- What are the origin-destination travel patterns, as determined from SmartRider and other available data, and how do they compare to modelled patterns?
- Are there any new patterns of travel which can be discovered through data mining of SmartRider data?

The purpose of this report is twofold, first, to present the results of the outputs of Project 4.1, and second, to provide a level of technical detail related to the implementation of the project.

Project 4.1 Scope

Project 4.1 comprises three case studies, two of which were developed to analyse underlying patterns in the SmartRider data, while the third case study focusing on visualisation and model interaction.

The first SmartRider case study analysed the relationship between public transport usage, and several established factors that influence travel behaviour, accounting for spatial and temporal patterns. Aspects of the analysis include identifying important factors relating to travel behaviour in a spatial context and identifying outlier regions, and regions that exhibit both slightly better, and slightly worse, public transport utilisation than that expected based on the underlying assumptions inherent in the analysis. The second case study centred on the application of "big data" concepts to the SmartRider data in order to derive public transport origin-destination information, which represents a non-trivial query over the dataset. This involved the development of methods to store the SmartRider data, and subsequently generate origin-destination matrices for given subsets of the data. The objective for this case study was to enable the Department of Transport to make comparisons with Origin-Destination data generated from existing models. Both studies make use of the journey origins within the SmartRider dataset, which is defined as the location of the public transport journey start, corresponding with an initial SmartRider tag-on.

A further case study was proposed to investigate methods for the presentation of model outputs and the visual exploration of the SmartRider, and associated datasets; this case study was designed to demonstrate the flexibility of the underlying software architecture proposed for the project, which acts as a pre-cursor, or prototype, for a decision support system for transport.

Framework

The project was implemented with a view to bridging the gap between data and model outputs. To achieve this, the analysis was conducted within a flexible *spatial data analytics* software architecture, outlined in the original proposal. Implementing the methodology with respect to the architecture comprises several stages: data acquisition, data management, exposing the data in a flexible manner (data access), data analysis, visualisation, and finally presentation of data within an interactive web dashboard environment. The use of a web-based presentation method, along with the proposed data access mechanism, enables multiple users to access a single data source, representing the point of truth data, along with associated model. The analytics architecture incorporates linked data with a spatial context; including the determination of associated indicators and dimensions derived from the linked data. This enables the automated generation of a model given a choice of parameters, thus linking the model and the data. Model can then be stored and made accessible via the web through a number of mechanisms, including a model summary interface, and the model outputs, enabling the model results to be incorporated into a web dashboard.

Data Acquisition, Management, and Access

The initial SmartRider data extract was received in February 2017, covering all Smart Rider tag-on and tag-off events over a period of approximately 17 months (from the 1st April, ending 30st September 2016). Updates were received monthly on data dated 3 months in arrears. This data was loaded into a relational database and exploration of optimising the storage method with respect to analysis was undertaken, with an SQL schema identified, in conjunction with a Document database approach.

The dimensions influencing travel behaviour were identified, and indicators that encapsulate properties of the dimensions were subsequently determined. Corresponding datasets for each indicator were then acquired from a variety of sources and incorporated into the analysis system. Sources included Landgate, the ABS Table Builder, and via ABS consultancy, which was require to obtain the Estimated Resident Population for 2015.

Data Analysis and Visualisation

The travel behaviour analysis case study examines the usage of public transport with respect to supply and demand characteristics related to public transport, incorporating aspects of both generators and attractors. The underlying analysis of the case study lies in the generation of a mathematical model to determine an expected public transport usage for an area, in this case for a suburb based on the underlying assumptions of the model.

The travel behaviour model extends previous work on spatiotemporal public transport usage patterns [1]. The mathematical model was developed to encapsulate the relationship between public transport usage and a number of established factors that influence travel behaviour. The public transport usage in this model is indicated by patronage corresponding to public transport journey origins. The influencing factors considered include several different dimensions, and the associated indicators that can be used to encapsulate the properties of each dimension, and included factors such as public transport service and usage provisions, socio-economic factors, employment, and population characteristics

The software architecture facilitated the model generation process. The subsequent issue centred on how to consume, or interact with, the travel behaviour model outputs. While there are a number of potential use cases, the one that was chosen as a focus for the interaction with the model outputs was the examination of the model residuals for the suburbs. The residuals in this case corresponds to difference between the observed public transport usage and that expected by the model. It is a property of the model that the residuals correspond to an indication of the utilisation of public transport within the suburb, with values below 0 displaying a lower utilisation than expected, and above 0 equating to a better utilisation of public transport than expected (given the assumptions associated with the model). Thus, the utilisation value represents a meaningful method for interaction with the model outputs via spatial and graph based visualisations, in particular, when the visualisations are linked, and interactive. Aspects of the analysis include identifying important factors relating to travel behaviour in a spatial context and identifying outlier regions, and regions that exhibit both slightly better, and slightly worse, public transport utilisation than that expected based on the underlying assumptions inherent in the analysis, This is lead to the exploration of a number of different methods for visualising, and interacting with, both the data and the analysis outputs

The web accessible data and models were presented using four dashboards, each corresponding to a different concept: exploration of the model results, exploration of the model results for a particular region, exploration of the data behind the model, and finally, exploration of the SmartRider data summarised at the stop level (comprising bus stops and train stations) — the data behind the data used to derive the model. In addition, a number of visualisations summarising aspects of the SmartRider data were also explored, with a selection presented in this report. These visualisations represent outputs from the exploratory data analysis phase of data analytics, which comprises using visualisations to explore properties of the underlying data.

The public transport origin destination (O/D) matrix case study experimented with a number of different data storage and handling techniques in order to determine a flexible method for the extraction of O/D matrices from the SmartRider data. A method was developed using a document database approach that enabled O/D matrices to be generated between spatial regions (such as STEM zones and suburbs, for a given set of date ranges. Further, the method can also incorporate filters based on SmartRider attributes, for example, to generate an O/D matrix for tertiary students. This approach was extended to incorporate destination locations, with an O/D matrix being generated based on the stops around a locations as a destination, with the remaining stops acting as origins.