Journal of Air Transport Management 49 (2015) 35–45

Contents lists available at ScienceDirect

Journal of Air Transport Management

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

Processing passengers efficiently: An analysis of airport processing times for international passengers

Jegar Pitchforth*, Paul Wu, Clinton Fookes, Kerrie Mengersen

Queensland University of Technology, Australia

A R T I C L E   I N F O

Article history:
Received 11 August 2014
Received in revised form
28 June 2015
Accepted 29 June 2015
Available online 14 August 2015

Keywords:
Hierarchical regression
Airport
Regression tree
Inbound passenger processing
Bayesian learning

A B S T R A C T

With immense and growing pressure on stakeholders in international airport terminals to process passengers faster than previously, there is a great benefit to understanding which factors affect passenger processing times and in which situations. In addition, storing and analysing the collected data in batch is itself a difficult and time consuming task that could be made much simpler with sequential analysis. We aim to present a method for airport managers to discover which variables are important to understanding passenger processing times and identifying problematic passenger profiles without the need for high computational capacity and full historical datasets.

In this paper we introduce Bayesian hierarchical models as a method of sequentially processing data, reducing computation time and obviating storage of large amounts of raw data. We use a range of exploratory models to identify which variables are important to predicting passenger processing time using a dataset from a day of operations at an international airport terminal, then compare a range of regression models. A Bayesian hierarchical regression model based on the model of best fit discovered through exploration is then applied to two subsets of data. We demonstrate that sequential updating based on daily data achieves similar results to batch processing based on full historical datasets and can therefore be used as an alternative in appropriate circumstances. Using the presented models, we find that the airline operating a flight is the most important variable to determining passenger processing time, followed by each passenger's age, sex and nationality. We demonstrate that in our dataset, the passenger profiles correlated with higher mean processing times overall were not the same as those passengers most problematic for meeting processing time targets.

© 2015 Elsevier Ltd. All rights reserved.

1. Introduction

The number of passengers moving through Australias airports puts significant pressure on the systems designed to record and facilitate their entry into the country, while also managing the risk of allowing entry to unapproved travellers. With funding being decreased to maintain these systems, and more passengers from a more diverse range of countries using Australian airports each year (Curran, 2012; Pan and Laws, 2003), it is of critical importance to understand the factors involved in determining the performance of these systems and to develop more efficient methods of analysis and evaluation.

The time taken to process passengers through immigration checkpoints after landing is an additive sum of the time taken for the passenger to conduct all mandatory processes (such as disembarking the aircraft) and discretionary activities (such as buying products in the Duty Free store) (Pitchforth et al., 2014). However, the performance of each sub-process cannot be determined directly from readily taken measurements; only the aircraft landing time and time of the passenger being processed are measured such that only overall passenger processing time is directly observed. This is a problem as individual targets are specified for each stakeholder, but different stakeholders are responsible for different sub-processes of inbound passenger facilitation (Wu et al., 2014). For example airlines are responsible for unloading aircraft, immigration checkpoints are the responsibility of the Australian Customs and Border Protection Service (ACBPS), and the search of luggage (which is not considered further in this paper) is the shared responsibility of the Department of Agriculture, Fisheries and Forestry (DAFF) and ACBPS. All these organisations work together in the airport to

* Corresponding author. Present/permanent address: Queensland University of Technology, Lvl 4 Y Block, Gardens Point Campus, Brisbane, Queensland 4000, Australia.

E-mail address: jegar.pitchforth@qut.edu.au (J. Pitchforth).

http://dx.doi.org/10.1016/j.jairtraman.2015.06.016
0969-6997/© 2015 Elsevier Ltd. All rights reserved.

facilitate the efficient and secure entry of passengers from overseas airports.

There are, however a number of variables recorded about the passengers arriving at the terminal and the flights from which they arrived. These variables systematically affect the time taken for the passenger to move through all these sub-processes, so understanding these variables is useful for predicting overall processing time. This allows managers of the system to adopt effective strategies for reducing predicted processing times below the required targets and account for variation in expected processing time, so that probabilities pertaining to individuals can be better estimated.

Another issue is that the operations of the airport generate a huge amount of data on a daily basis which are difficult to store in the long term. Large datasets are also very difficult to process, often requiring specialised hardware to cope with the computational requirements for processing the high number of entries. In current industry practice data are required to be batch processed (i.e. using full historical datasets (Ikura and Gimple, 1986)), so this problem becomes more difficult to address as time goes on and the data accumulates, generating larger and larger datasets. Very quickly the computational requirements for such large datasets become intractable. However, there are sequential methods available that do not require a complete historical dataset for useful predictions. Modelling the data with sequential methods would allow managers to use model outputs for supporting their decisions without the need for vary large historical datasets, obviating complicated data storage and access arrangements.

Passenger flow modelling has been explored in the past using a range of methods. The earliest attempts were based on queuing models focussed on passengers' interaction with baggage delivery (Tanner, 1966), security (Gilliam, 1979) or overall terminal design (McKelvey, 1988). While the interest in this approach has waned in recent times, there are examples of sophisticated recent work using this method (Cochran and Roche, 2009). Another popular and well supported approach to modelling passenger flows in airports recently has been to use simulation (Mumayiz, 1990; James, 2009). Simulation models have been applied to passenger movements with great success (Paullin, 1966; Ma, 2013; Ray and Claramunt, 2003; Kovacs et al., 2012). In particular, Agent based modelling (Bonabeau, 2002) is a very common type of simulation model for airport management, and as such has been applied to data from a range of terminals (Pendergraft et al., 2004; Jim and Chang, 1998; Gatersleben et al., 1999). However, while this approach gives an impression of the overall behaviour of the terminal based on known parameters, it neither gives managers insight into how passenger characteristics are reflected in the data nor does it give any estimation of the uncertainty around model predictions. Process models, are often used to provide an overall view of terminal processes with regards to processing times and passenger capacity (Andreatta et al., 1999; Brunetta et al., 1999; Gatersleben et al., 1999; Henderson, 1974). These models use deterministic formulae that each represent specific parts of the airport terminal process. While this approach is a useful step in modelling complicated systems, it provides too broad an interpretation of system behaviour to be used in meaningful decision support as there is no way to quantify uncertainty around model predictions.

This paper analyses passenger arrival and flight log data to determine which variables are important to predicting performance against set targets and passenger processing time, as well as selecting a model that best fits the data. First we seek to find which set of variables provides the best fit to the data using exploratory methods. Second, we demonstrate that a Bayesian hierarchical model based on the variables discovered in the first step can be used to process data sequentially. In doing so we demonstrate the application of a method novel to passenger processing literature by which managers can analyse their own airport's data for use in decision support.

In situations where there is very little known about the dataset, a range of reasonable models must be selected to determine which set of variables explains the greatest proportion of variance in the outcome, without adding unnecessary model complexity through variables with low explanatory power. The possible models arising from the exploratory analysis are compared to determine the model with best fit to the data. This approach has been used extensively in the past in situations where very little is known about the importance and structure of variables in the dataset (McArdle and Ritschard, 2013), but has not yet been applied to airport terminal data.

In this paper we use four approaches to identify important variables for predicting passenger processing time and determine a preferred model through which predictions can be made. In the first instance a boosted regression tree is run to determine which variables have the most relative influence on the processing time. These models are based on model-averaging algorithms, combining the results of many regression trees to arrive at an additive regression model describing the relative influence of each factor on the output variable (Elith et al., 2008). Once we have an overall understanding of the variables being considered conditional inference trees are generated to determine which levels of which variables partition the population most effectively with regards to processing time. Conditional inference trees, like regression trees seek at each step to divide data as evenly as possible based on levels of included factors (Hothorn et al., 2006), arriving at a group classification based on binary (in the case of classification trees) or continuous (in the case of regression trees) outcome values. However, conditional inference trees provide an unbiased approach more fitted to scenarios where model factors have differing numbers of levels.

Once the variables of interest have been established through these techniques, we employ linear regression and multilevel regression models to gain a better understanding of which set of variables explains the greatest amount of variance in the data without including variables with low explanatory power, as well as the most useful structure for the variables. Both of these approaches are cases of Generalized Linear Modelling, which each rely on different assumptions. With linear regression we attempt to fit all the variables around a single regression line, implying that we assume all model factors are reflecting the same trend (Seber and Lee, 2012). However, with multilevel regression we can assume that one set of variables varies randomly, with systematic, or fixed, effects of variables within that grouping (Austin et al., 2000). For example, processing times for passengers might vary depending on their age and nationality, but the effect and importance of these variables may be dependant on the flight they were on. Multilevel regression allows us to acknowledge in our model that passengers from the same flight are likely to be more similar to each other than to passengers from other flights, and control for this similarity in our final analysis.

A Bayesian approach is then adopted because it not only provides the required predictions, but can also naturally facilitate the requirement of sequential updating, also known as Bayesian learning. The model can also encode the relationships in the data with the capability of simulating statistically similar data when needed, obviating storage of raw datasets. Another advantage of this is that it reduces overall computation time by relying on prior parameters rather than full datasets. To our knowledge, this is the first time Bayesian hierarchical regression models have been proposed for assessing passenger processing time in airports as well as one of the first in-depth explorations of real passenger processing data from an international airport. We describe this type of model

Download English Version:

https://daneshyari.com/en/article/1030720

Download Persian Version:

https://daneshyari.com/article/1030720

Daneshyari.com