



ELSEVIER

Expert Systems with Applications 28 (2005) 105–117

Expert Systems
with Applications

www.elsevier.com/locate/eswa

The development of an adapted Markov chain modelling heuristic and simulation framework in the context of transportation research

Davy Janssens*, Geert Wets, Tom Brijs, Koen Vanhoof

Limburgs Universitair Centrum, Data Analysis and Modelling Group, Transportation Research Institute, Universitaire Campus, Gebouw D, B-3590 Diepenbeek, Belgium

Abstract

This paper has developed and evaluated the implementation of an adapted Markov Chain modelling heuristic and simulation framework in the context of transportation research. In order to gain insight into the travel patterns of individuals and in the decision making that people use to make transport mode decisions, a new methodology is presented in this paper to extract knowledge from data. The presented approach shows two ways to store the sequential information (sequences of activities and travel) that is typically incorporated in activity diary data. The approach is novel, especially with respect to store information in ‘codebooks’, a term which is introduced to reflect that the information which is kept, represents the combinations of activities that typically sequentially occur in a persons’ diary.

In order to test the validity of the heuristic, new data is simulated and compared with the original observed data. The new data is generated by means of Monte Carlo simulation and the empirically derived information from the codebooks is used as a constraint in the simulations. In order to make a mature evaluation of the simulated diaries, different performance indicators were considered by using pattern-, trip- and activity-level measures. It is shown in the paper that the results are satisfactory and that the framework that was developed holds out considerable promise; both for gaining behavioural decision making insights and for simulating activity diary data that can assist practitioners and researchers in the calibration of travel demand models.

© 2004 Elsevier Ltd. All rights reserved.

Keywords: Markov Chain modelling; Transportation research; Monte Carlo simulation

1. Introduction

For the last decade, activity-based transportation models have set the standard for modelling travel demand. The most important characteristic in these models is that travel demand is derived from the activities that individuals and households need or wish to perform. The main advantage is that travel has no longer an isolated existence in these models, but it is perceived as a way to perform activities and to realize particular goals in life. The aim of this paper is not to detail on these models but it is important to realize that the increased complexity of an activity-based transportation model is also reflected in the data that are required to estimate such a model. Indeed, the collected travel

information needs to be immediately associated with the activities the respondent says to be engaged in. This implies that data requisites are considerably higher. Fortunately, this wealth of information also has the potential to obtain a better understanding about the behavioural mechanisms and principles that individuals and households use to organize activities and to perform travel activities.

The aim of this paper is therefore to present an adapted Markov Chain modelling heuristic that can satisfy both needs; i.e. it should both be possible to (i) capture and increase insights in behavioural decision making and (ii) provide a simulation tool of activity diary data that can assist practitioners and researchers in the calibration of activity-based transportation models. The work in this paper is a thorough extension of previously published work (Janssens, Wets, Brijs, & Vanhoof, 2004). Other approaches that deal with scarcity of data and that provide insights in behavioural decision making are mentioned in Arentze, Katoshevski, and Timmermans (2001), Greaves, and Stopher (2000)

* Corresponding author. Tel.: +32 0 11268643 fax: +32 0 11268700.

E-mail addresses: davy.janssens@luc.ac.be (D. Janssens), geert.wets@luc.ac.be (G. Wets), tom.brijs@luc.ac.be (T. Brijs), koen.vanhoof@luc.ac.be (K. Vanhoof).

and Stopher, Greaves, and Bullock (2003). The idea to come up with an adapted Markov Chain modelling heuristic was advanced after it became clear that the simple application of the available Markov Chain modelling theory is infeasible in this application area.

Typically, data are collected by means of activity diaries after it has been repeatedly shown (Koppelman, 1981; Robinson, 1985) that traditional travel surveys especially under-report off-peak, non-home based trips of short duration. Stopher (1992), Clarke, Dix, and Jones (1981) and Niemi (1993) have argued that activity diaries outperform travel surveys in this respect. Moreover, it is claimed (Clarke et al., 1981) that despite the fact that only out-of-home activities generate traffic, collecting in-home activities seems also warranted since it provides the closest correspondence with the natural storing of information and the planning of activities.

It is assumed in this paper that each diary consists of a set of correlated successive observations of a random variable. To this end, a discrete random variable X_t is considered, taking values in the finite set $\{1, \dots, m\}$, where each value in this set represents an activity that occurs in a persons' diary. Travelling is considered as an activity as well, however, the transport mode is added as an additional attribute in this case. Our goal is to simulate (predict) the value taken by X_t as a function of the values taken by previous observations of this variable. Markov chains are probabilistic models which are commonly used to model this type of dependencies in data. However, as said, their application resulted in additional difficulties in this research context (see *infra*). For this reason, the development of a new adapted Markov Chain modelling heuristic was implemented.

The presented approach shows two different ways to store the sequential information (sequences of activities) in 'codebooks', a term which is introduced to reflect that the information which is kept here represents low- and high-order combinations of activities that typically sequentially occur in one particular diary. In the limit, low-order combinations assume that the present value taken by X_t is entirely explained by the first lag (Activity $t-1$), while high-order combinations assume that the correct present value of X_t can only be explained by the last $k-1$ observations (Activity $t-1$, Activity $t-2$, ... Activity $k-1$) in which k represents the length of the diary. However, high-order combinations tend to overfit the sample dataset, while low-order combinations have the problem that the knowledge extraction is insufficiently tuned towards the sample data and that therefore insights in activity-travel patterns are non-reliable. A trade-off will be explored in the paper. Hereafter, new data is simulated based on the sequential information, which is incorporated in the codebooks. The new simulated data is then compared with the originally observed data. The aim of this simulation framework is two-fold: (i) it serves as a validation measure for the developed heuristic and

(ii) it is a simulation tool of activity diary data that can assist practitioners and researchers in the calibration of activity-based transportation models.

The remainder of the paper is organized as follows. Section 2 briefly introduces Markov Chains in order to provide the necessary theoretical background. Section 3 elaborates on the problems that were encountered when using Markov chains and it describes the reasons for developing the new heuristic. In Section 4, the core of the algorithm is introduced. Section 5 describes the procedure which is used for simulating the new data based on the information which is stored in the codebooks. A controlled Monte Carlo simulation technique is introduced. In Section 6, the developed method is tested and empirically validated by means of pattern-, trip- and activity-level performance indicators. Conclusions and topics for future research are given in Section 7.

2. Background and previous work: Markov chain modelling

Capturing dependencies between successive observations of a random variable is mainly done by means of Markov Chains until now. Markov chains are probabilistic models, which were introduced by Andrej Andreevic Markov at the beginning of the 20th century. Their application domains have been numerous, including geography, biology, meteorology, music, and many others. The fundamentals of Markov chain modelling are introduced in this section.

Consider a discrete-time random variable X_t , taking values in the finite state space $\{1, \dots, m\}$. In order to predict the value taken by X_t , the Markov property is used as a necessary condition in Markov chain modelling. The Markov property says that the present time t , is entirely explained by the first lag ($t-1$), so that we can write:

$$\begin{aligned} P(X_t = i_0 | X_0 = i_t, \dots, X_{t-1} = i_1) \\ = P(X_t = i_0 | X_{t-1} = i_1) = q_{i_1 i_0}(t), \end{aligned}$$

where $i_t, \dots, i_0 \in \{1, \dots, m\}$. (1)

Considering all combinations of i_1 and i_0 , we can now construct a transition matrix Q , each of whose rows sums to 1.

$$Q = \begin{array}{c} X_t \\ X_{t-1} \end{array} \begin{array}{c} 1 \quad \dots \quad \dots \quad m \\ \left| \begin{array}{cccc} q_{11} & \dots & \dots & q_{1m} \\ \vdots & \ddots & & \vdots \\ \vdots & & \ddots & \vdots \\ m & q_{m1} & \dots & q_{mm} \end{array} \right| \end{array}$$

متن کامل مقاله

دریافت فوری ←

ISIArticles

مرجع مقالات تخصصی ایران

- ✓ امکان دانلود نسخه تمام متن مقالات انگلیسی
- ✓ امکان دانلود نسخه ترجمه شده مقالات
- ✓ پذیرش سفارش ترجمه تخصصی
- ✓ امکان جستجو در آرشیو جامعی از صدها موضوع و هزاران مقاله
- ✓ امکان دانلود رایگان ۲ صفحه اول هر مقاله
- ✓ امکان پرداخت اینترنتی با کلیه کارت های عضو شتاب
- ✓ دانلود فوری مقاله پس از پرداخت آنلاین
- ✓ پشتیبانی کامل خرید با بهره مندی از سیستم هوشمند رهگیری سفارشات