Transportation Research Part D 47 (2016) 54–66

Contents lists available at ScienceDirect

Transportation Research Part D

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

Modeling time-of-day car use behavior: A Bayesian network approach

Dawei Li [a,*], Tomio Miwa [b], Takayuki Morikawa [c]

[a] Jiangsu Key Laboratory of Urban ITS, Southeast University, Jiangsu Province Collaborative Innovation Center of Modern Urban Traffic, Sipailou 2, Xuanwu District, Nanjing 210096, China
[b] EcoTopia Science Institute & Green Mobility Collaborative Research Center, Nagoya University, Japan
[c] Graduate School of Environmental Studies & Green Mobility Collaborative Research Center, Nagoya University, Japan

ARTICLE INFO

Article history:

Keywords:
Car use
Bayesian networks
Latent class
Machine learning
GPS data

ABSTRACT

In this research, a Bayesian network (BN) approach is proposed to model the car use behavior of drivers by time of day and to analyze its relationship with driver and car characteristics. The proposed BN model can be categorized as a tree-augmented naive (TAN) Bayesian network. A latent class variable is included in this model to describe the unobserved heterogeneity of drivers. Both the structure and the parameters are learned from the dataset, which is extracted from GPS data collected in Toyota City, Japan. Based on inferences and evidence sensitivity analysis using the estimated TAN model, the effects of each single observed characteristic on car use measures are tested and found to be significant. The features of each category of the latent class are also analyzed. By testing the effect of each car use measure on every other measure, it is found that the correlations between car use measures are significant and should be considered in modeling car use behavior.

Introduction

The use of private cars is a major cause of congestion and air pollution (Goodwin, 1996). The modeling of car use behavior is essential if we are to measure the environmental and social costs of urban transportation systems, evaluate the effect of certain transportation policies (e.g. road pricing) on reduced car use, and predict travel demand.

Car use behavior depends on many factors. There is a heavy dependence on the characteristics of a household's principal driver, such as age, gender, and occupation. Previous research has shown that workers, young people, and males are likely to drive more (Hensher, 1985; Mannering, 1983; Train, 1986). Vehicle characteristics also affect car use behavior. For example, Van Wissen and Golob (1992) determined the relationship between car usage and choice of fuel type. Household characteristics, such as household size and housing quality, are another category of factors that affect car use patterns (Borgoni et al., 2002). At a more macroscopic level, land use and transport policies also affect car use behavior (Gärling et al., 2002; Jakobsson et al., 2002; Kitamura et al., 1997). From a psychological and behavioral perspective, car use patterns are also affected by attitudes, motives, and habits (Gärling et al., 1998; Gardner and Abraham, 2008; Steg, 2005).

In all of the earlier research referred to above, the measures of car usage were not time-dependent. They included such indexes as annual mileage and weekly car use frequency. However, the modeling of dynamic behavior is the inevitable trend

* Corresponding author. Tel.: +86 25 8379 5642.
E-mail addresses: lidawei@seu.edu.cn (D. Li), miwa@civil.nagoya-u.ac.jp (T. Miwa), morikawa@nagoya-u.jp (T. Morikawa).

http://dx.doi.org/10.1016/j.trd.2016.04.011
1361-9209/

in transportation research. A statistical car use measure such as annual mileage can only roughly reflect the car use behavior of drivers; drivers with similar values of statistical car use measures may have different car use pattern.

The modeling of time-of-day car use behavior is also required for some particular applications. As an example, Harris and Webber (2012) did a statistical analysis on time-of-day car use patterns and analyzed their impact on the provision of vehicle-to-grid services. In another investigation related to electric vehicles, describing drivers' time-of-day car use was necessary to assess the effect of battery-range limitations on the popularization of plug-in electric vehicles (Pearre et al., 2011). In these two investigations, as well as in analysis by Krumm (2012), time-of-day car use behavior was described using only basic statistics; relationships with driver or other characteristics were not modeled. Full modeling of time-of-day car use behavior has not been described in the literature, according to our limited investigations. Therefore, the main contribution of this research is to model drivers' time-of-day car use behavior and analyze its relationship to driver characteristics and other characteristics.

From the perspective of the modeling techniques used, previous research either used a single car use measure, or estimated separate models for each of multiple car use measures. However, the correlations between various car use measures are obvious: people who use their cars more frequently are more likely to accrue higher annual mileage. Furthermore, in a time-of-day analysis, each car use measure will be calculated for multiple time intervals. Therefore, in this research, the correlations among the various car use measures will be considered. Consequently, besides modeling the relationships between car use measures and observed characteristics, the main methodology challenge of this research is the representation of the correlations among multiple car use measures in different time intervals.

To this end, in modeling the complex relationships among multiple car use measures in different time intervals, we eschew the econometric models used in most travel behavior research. Instead we choose to use a Bayesian network (BN), a commonly applied model in the field of machine learning. The main advantages of the BN approach are explained in the following section.

This paper is organized as follows. Section 'Bayesian networks' presents some background relating to our modeling technique using a BN. Section 'Data description and basic statistics' describes the data used in this research and gives some basic statistics. Section 'BN model of time-of-day car use behavior' develops the BN model of time-of-day car use behavior. Section 'Inference and analysis' analyzes drivers' car use behavior according to inferences obtained from the learned BN. Finally, Section 'Conclusions' presents the conclusions of the research.

Bayesian networks

In this section, we describe some of the basic characteristics of Bayesian networks (BNs). Also called belief networks, Bayesian belief networks, Bayes nets, and sometimes also causal probabilistic networks, BNs are an increasingly popular method of modeling uncertain and complex relationships among multiple variables. As a popular tool for machine learning, they have also been widely used to solve practical problems in the transportation field, such as the development of Intelligent Transportation Systems (Li et al., 2011; Ozbay and Noyan, 2006; Zhang and Taylor, 2006), travel behavior analysis (Janssens et al., 2006; Li et al., 2013), and travel demand modeling (Castillo et al., 2008, 2012).

Bayesian networks are directed acyclic graphs that allow efficient and effective representation of a joint probability distribution over a set of random variables. Formally, a Bayesian network for a set of random variables $\mathbf{X} = \{X_1, \ldots, X_n\}$ is the pair $B = (G, \boldsymbol{\theta})$. The first component, G, is a directed acyclic graph whose nodes correspond to the random variables $X_1, \ldots, X_n$ and whose links represent direct dependencies between the variables. Graph G, which is also called the structure of this Bayesian network, encodes the independence assumption: that each variable X_i is independent of its non-descendants given its parents in G. The second component of the pair, $\boldsymbol{\theta}$, represents the set of parameters that quantifies the network. It contains a parameter $\theta_{X_i|\Pi_{X_i}} = P_B(X_i|\Pi_{X_i})$ for each possible value x_i of X_i, and Π_{x_i} of Π_{X_i}, where Π_{X_i} denotes the set of parents of X_i in G. A Bayesian network B defines a unique joint probability distribution over $\mathbf{X}$ given by

$$P_B(X_1, \ldots, X_n) = \prod_{i=1}^{n} P_B(X_i|\Pi_{X_i}) = \prod_{i=1}^{n} \theta_{X_i|\Pi_{X_i}} \tag{1}$$

The reasons for using BNs in this research relate to the main advantages of Bayesian networks from the perspective of application (Uusitalo, 2007), which can be outlined as:

(1) *Suitable for small data sets*: there are no minimum sample sizes required to estimate a BN model. It has been demonstrated that Bayesian networks offer good prediction accuracy even with rather small sample sizes (Kontkanen et al., 1997).
(2) *Capable of dealing with missing data and incorporating latent variables*: the parameters of the model can be estimated from incomplete data using an Expectation-Maximization (EM) algorithm (Lauritzen, 1995; Spiegelhalter et al., 1993), a technique that will be used in this research. As a special case of missing data, latent variables can also be incorporated in a BN.
(3) *Explicit uncertain inferences*: the relationship between actions, knowledge and uncertainty can be considered explicit in the structure and parameters in a BN (Jensen and Nielsen, 2007). The updating of probabilities when some evidence is obtained for certain variables can be very easily done using certain inference algorithms.