Available online at www.sciencedirect.com

ScienceDirect

Transportation Research Procedia 26 (2017) 53–61

Transportation Research Procedia

www.elsevier.com/locate/procedia

44th European Transport Conference 2016, ETC 2016, 5-7 October 2016, Barcelona, Spain

A dynamic Bayesian network approach to forecast short-term urban rail passenger flows with incomplete data

Jérémy Roos [a,b,*], Gérald Gavin [b], Stéphane Bonnevay [b]

[a] RATP, 75012 Paris, France
[b] Université de Lyon, ERIC EA 3083, 69100 Villeurbanne, France

Abstract

We propose an approach to forecast the short-term passenger flows of the urban rail network of Paris. Based on dynamic Bayesian networks, this approach is designed to perform even in case of incomplete data. The structure of the model is built so that the flows are predicted from their spatiotemporal neighbourhood, while the local conditional distributions are described as linear Gaussians. A first application carried out on a single station highlights the need to incorporate information on the transport service. In the presence of missing data, we perform the structural expectation-maximization (EM) algorithm to learn both the structure and the parameters of the model. Short-term forecasting is conducted by inference via the bootstrap filter. Finally, we apply the model to an entire Paris metro line, using on-board counting, ticket validation and transport service data. The overall forecasting results outperform the historical average and last observation carried forward (LOCF) methods. They also evidence the key role of the transport service in the modeling.

Peer-review under responsibility of the Association for European Transport

Keywords: short-term passenger flow forecasting; urban rail network; dynamic Bayesian networks; incomplete data.

1. Introduction

RATP is the main public transport operator in the Paris region. It operates 16 metro lines, sections of 2 RER (commuter rail), 8 tramway lines and more than 350 bus lines. Currently, it uses several tools for passenger flow modeling, including GLOBAL and IMPACT (Leblond and Garcia Castello, 2016). These models are mainly

* Corresponding author.
E-mail address: jeremy.roos@ratp.fr

2352-1465
Peer-review under responsibility of Association for European Transport
10.1016/j.trpro.2017.07.008

designed to assess the effects of infrastructure or transport policy changes over the long term. Hence, they do not provide short-term predictions and cannot take into account the effects of unanticipated or non-recurring events, such as service disruptions, closures of stations or crowd-attracting events. On the other hand, the diversity of available data is still largely untapped. Each service often works with a limited number of data sources and only has a partial view of the passenger mobility within the network. Moreover, these services have to face with incomplete data caused by failures or lack of collection systems. This issue leads to the implementation of imputation methods which can be time-consuming.

From these observations, we propose a model to forecast the short-term passenger flows of the RATP urban rail network (metro and RER). The chosen approach is based on dynamic Bayesian networks, which are able to combine heterogeneous information sources and make predictions even in case of missing data. The model is designed to cater for various applications in transport system management, such as operation planning, passenger flow regulation or passenger information.

In this paper, we first present a brief state of the art of short-term traffic forecasting. After introducing the Bayesian network representation, we perform a quick experiment on a single RER station with a complete set of ticket validation and counting data. Then we extend the model to the time factor, using dynamic Bayesian networks, and incorporate information on the transport service. The learning and inference processes are detailed for incomplete data and the application area is extended to an entire metro line. Finally, we conclude the paper and discuss the limitations of our approach.

2. Short-term traffic forecasting: brief state of the art

There is a vast scientific literature on short-term traffic forecasting. Various methods have been explored, which can be classified into naïve, parametric and nonparametric methods (van Hinsbergen et al., 2007). Due to their easy implementation, naïve methods have been widely used, including historical average (Smith and Demetsky, 1997) and last observation carried forward (LOCF), also called random walk (Williams, 1999). Among the parametric methods, ARIMA models have been explored since the late 1970's (Ahmed and Cook, 1979). A great focus has also been given to Kalman filter algorithms (Okutani and Stephanedes, 1984). In recent years, nonparametric methods such as nonparametric regression (Clark, 2003) and neural networks (Vlahogianni et al., 2005) have been successfully applied. Their ability to better model nonlinear processes has contributed to their popularity.

In his thesis, Haworth (2014) points out the lack of consensus on the most suitable method. Indeed, each dataset has its own characteristics (e.g., spatiotemporal resolution, transport network, type of data), which makes comparison to other studies difficult. Moreover, for time-saving purposes, the authors generally compare their sophisticated model to basic implementations of other models.

Missing data is a common problem in real-world situations, including traffic flow forecasting. Surprisingly, few authors have dealt with this problem in a real-time setting (Haworth and Cheng, 2012). Hence, most of the short-term forecasting models are ill-equipped to provide real-time predictions in case of incomplete data. Among the few solutions that have been proposed, particular attention has been afforded to Bayesian networks, both on urban road networks (Sun et al., 2006) and on highways (Whitlock and Queen, 2000).

Most research in short-term traffic forecasting has focused on vehicle flows in road networks. By contrast, little work has been devoted to passenger flows in public transport networks, the existing models being mainly designed for long-term planning (Ma et al., 2014). Some authors have recently begun to tackle this issue, especially through neural network approaches (Celikoglu and Cigizoglu, 2007; Wei and Chen, 2012; Li et al., 2013). In these models, the flows are generally predicted on the basis of their historical values. The input data may also include other features, such as information from adjacent flows (Li et al., 2013), temporal factors (Wei and Chen, 2012; Li et al., 2013) or weather conditions (Li et al., 2013). Although the service provided by the public transport operator has a potential impact on the passenger flows, it seems that this information has not yet been considered.