Machine learning and data mining for urban mobility intelligence

Etienne Côme¹, Latifa Oukhellou¹, Allou Sameé¹ & Lijun Sun²

¹- Univ. Gustave Eiffel, COSYS, GRETTIA
F-77447 Marne-la-Vallée, France
²- McGill University,
817 Sherbrooke St. W. Montreal, QC H3A 0C3, Canada

Abstract. The last few decades have seen a faster development of digital systems for observing the mobility of people and goods. Various sensing systems - such as radio communication, Wi-Fi, Bluetooth, validation of smart cards, mobile phone, and road traffic monitoring systems - have enabled researchers and practitioners to acquire large amounts of data, which generally refer to individual and collective trajectories. The mobility data can be further enriched with side information, such as text corpora from social media, survey data, and weather information. These massive data, temporally and spatially structured, can benefit from advanced machine learning and data mining methods, providing decision aid tools, and contributing to the development of safer, cleaner, and more efficient transportation systems. They can also help to implement new mobility services for the user. This article provides an overview of methodological advances in temporal and spatial mobility data processing.

1 Introduction

This special session will be an opportunity to present recent advances in mobility data analysis. This field refers to descriptive methods, generally based on unsupervised learning, highlighting the main uses, routines, mobility patterns, and provides a basis for anomaly detection. The second field of interest in mobility data processing in the scope of this special session concerns predictive analytics, which performs in a supervised learning framework to predict quantities of interest such as transport demand or operating conditions. The implementation of these methods faces challenges related to the incompleteness, heterogeneity, and strong temporal and spatial correlation within the data, and their high dimensionality or volume.

The paper is organized as follows. Section 2 details structural time series decomposition applied to mobility data and gives an example of its application. Section 3 presents predictive models dedicated to forecast mobility demand. Section 4 focuses on tensor learning and factorization applied to mobility data.

2 Structural time series decomposition applied to mobility data

Considering mobility data as time series, structural models [1] have been used to extract from them multiple latent components, each representing an aspect of
the original profile [2, 3]. One of the advantages of such decompositions is that
the resulting components can be interpreted and predicted more easily than the
raw series. In particular, it becomes possible to identify specific patterns such as
a long-term trend, repeating seasonal patterns (day, week, year), calendar phe-
nomena (bank holidays ...), or the influence of exogenous phenomena (concerts,
strikes, health crisis ...). In the following, we first describe these models and
then show their practical use on mobility data. This work is part of the study
conducted in [3].

2.1 Structural time series modeling

Given an observed time series \((y_1, \ldots, y_T)\), with \(y_t \in \mathbb{R}\), the typical structural
model can be written as follows:

\[
y_t = \ell_t + s_t + \sum_{j=1}^{p} \beta_j^{(j)} x_t^{(j)} + \varepsilon_t, \quad \varepsilon_t \sim \mathcal{N}(0, \sigma^2),
\]

where \(\ell_t, s_t, \varepsilon_t\) are the trend, seasonal and noise components, and the \(\beta_j^{(j)}\)
are the coefficients associated to \(p\) known covariates \(x_t^{(j)} \in \mathbb{R}\). To model their dynamic
effects, the components \(\ell_t, s_t\) and \(\beta_j^{(j)}\) are usually represented in a stochastic
form using auto-regressive processes [1]. Structural models are a more flexible
alternative to iterative decomposition schemes such as Seasonal and Trend de-
composition using Loess (STL) [4], or Census X-13ARIMA [5]. Their advantage
lies in the fact that they are part of the more general framework of state-space
models [6] and thus can benefit from estimation tools such as the Kalman filter,
the Expectation-Maximization (EM) algorithm and Bayesian model selection
criteria. However, their practical implementation requires databases collected
over relatively long time periods and the incorporation of \textit{a priori} knowledge in
the model to increase their interpretive power. They have been widely used in
several application fields such as economics [7], tourism [8], meteorology [9] or
energy consumption [10] and have been found to be of great interest in analyzing
mobility data.

The case study presented in this article concerns the decomposition of daily
ridership data at a major station (La Defense Grande Arche: RER A) of the
Paris railway network, which allowed us to answer questions such as:

- How do the variations in the original time series which is very often noisy
  translate into each component?

- What is the impact of exogenous events on passengers’ decisions to use
  public transport?

The data were collected over a nine-year period, which constitutes an opportu-
nity for this type of analysis. Indeed, this relatively long period of data collection
allows the advent of several punctual or redundant events such as maintenance
works, strikes, COVID-19 health crisis. The structural model retained for these
data has a double seasonality (weekly and annual) and has been used with some regularity constraints on the trend and seasonal component dynamics [3]. Figure 1 shows the time series and its main resulting components. As it can be seen, a time series of public transport ridership is very noisy. Its decomposition into several components to which we assign meaning is therefore useful for a better analysis of long-term and mid-term phenomena on the ridership, such as trend and seasonal variations (annual, weekly), as shown in Figure 1.

![Figure 1: Decomposition of the time series ($y_t$) of flows entering a railway station (La Defense Grande Arche: RER A) into log-scaled trend ($\ell_t$), annual seasonality ($s_t$), and residuals ($\epsilon_t$); Figure taken from reference [3]](image)

The decomposition also allows a better understanding of the impact of specific exogenous events on passenger usage of public transport. Figure 2 shows for this purpose the regression component related to the Covid-19 crisis, and more particularly the lockdown and post-lockdown periods. The health crisis has strongly impacted people’s movements: most of the office workers who travel daily could not go to work during this period because of the lockdown. As it can be observed, the post-lockdown period also significantly impacted the public transport usage since telecommuting and social distancing was still recommended.

It should be noticed that in mobility analyses, structural decomposition models have also been used for prediction purposes. The authors in [2] developed structural models to predict the rate of a bike sharing system usage. We can also cite the work carried out by [11], where these models were used for short-term prediction of bicycle flows. In their study on the prediction of time series of car counts at some Dublin intersections, the authors in [12] presented the decomposition of a time series into three components: trend, seasonality, and residuals. In [13], the authors used the same kind of model to forecast the monthly traffic volume of the next twelve months from a key corridor in New-Jersey.

The presented use case also opens the perspective of simultaneously modeling several time series associated with multiple stations, as in [12]. This multivari-
ate approach could be extended to search for common dynamic factors within mobility time series.

3 Predictive models to forecast mobility demand

Predictive models and machine learnings techniques are used quite commonly to predict car traffic and have now diffused into main stream applications like google maps or waze. Such applications use floating car data at a massive scale for short term traffic forecasting. Other transportation modes also benefit from such works with other data sources. This section will detail some of the works done recently with smart card data and trains loads data in the context of public transport demand prediction and put them in perspectives with forecasting approaches used in the transportation field more generally.

Since 2004, the use of smart card data to analyse mobility in public transportation has received substantial attention from researchers. More recently, studies on mobility analysis have revolved around passenger demand forecasting. A distinction can be made between research that relates to forecasting Origin-Destination (OD) matrices and research that attempts to forecast passenger flows at a specific point or the load of a specific vehicle (e.g. train, bus). Knowledge about these factors is indeed essential for planning, operation and management in any transportation network, but each of these areas uses different types of data.

The passenger demand forecasting goals differ depending on the time horizon. Long-term forecasting aims to forecast demand with data available at the long-term period in advance (e.g., time features and planned events), which can be very useful for improving transport supply schedule. In contrast, the forecasting process can also account for the last observations, in which case it is generally referred to as short-term forecasting. In the future, in the case of an atypical situation, the main goal for transport operators is to use the forecasted passenger demand to optimize transport system operation to match transport supply to the atypical demand or propose to the passenger an alternative way to reach
their destination.

Short-term forecasting, which corresponds to a few time ahead forecasting, has been studied with different models. As an example, [14] used multiscale RBF networks to forecast the number of alighting passengers at different Beijing subway stations multiple time steps ahead (t+15 and t+30 minutes) by taking the number of boarding passengers at the other stations of the subway network into account. In this study, the authors performed an in-depth analysis of the results obtained under particular event scenarios. Other examples of subway passenger flow forecasting include the work of [15], where the authors predicted passenger flows of the next time step (t+2 minutes). The authors used a Bayesian network model and predicted multiple passenger flows (entry and exit) at all the stations of a subway line of the Paris network. In recent years, recurrent neural networks (RNN), in particular LSTM (introduced by [16]), have shown their efficiency in load prediction compared to standard neural networks by managing temporal dependencies [17, 18, 19, 20]. LSTM architectures are powerful in dealing with non-linear temporal dependencies in data. However, these models are usually combined with a spatial component to extract features from a network to address spatial dependencies. Thus hybrid models [19, 21, 22] divided the task into a temporal brick, most of the time an LSTM, and a spatial brick. The spatial brick can take many forms. For instance, a two-dimensional network [19] can consider spatio-temporal correlation in the traffic network. For multi-lane short-term traffic prediction, [21] replaced a standard LSTM with a convolutional LSTM to address large transport networks. Attention mechanisms (popularized by [23]) can be used in one-dimensional convolutional neural networks (1D CNN) and Gated Recurrent Units (GRU) to improve the generalization capacity of the models [22].

The above-mentioned studies are therefore built on regular time sampling, ignoring the real transport schedule. They forecast aggregated outflows and inflows over a fixed time interval for stations or metro lines. Temporal variability within the time series due to the transport schedule induces irregular temporal sampling. It makes it difficult to apply techniques that usually exploit the structural regularity of time series [18] such as previous LSTM or ARIMA frameworks. [18] used an encoder-predictor to take into account the transport schedule in their LSTM train load predictor. Another approach to deal with this specificity was introduced recently [24]. With such an approach, train load data are processed with an image-oriented methodology. First, the metro traffic is represented by generating a multi-channel image, exhibiting the spatial information of the trains running on a metro line while taking into account the irregular temporal sampling of the train loads. Input image denotes the ticket validation, time interval between trains, and waiting time at stations for all trains (in columns) and all station stops (in rows) over some time. Eventually, the output image represents the train loads. The prediction problem boils down to an inpainting / image-to-image translation problem that can be addressed with the standard deep-learning architecture used for image processing (e.g. U-net, CNN).
4 Tensor Learning and Factorization

A tensor is a multidimensional array [25], a natural way to organize urban mobility data. For example, we can represent the trip information of a metro system into a three-order (origin × destination × time) tensor. Each element in the tensor denotes the number of trips at a certain origin, destination, and time. Richer information, such as mode and user type, can be easily represented in a higher-order tensor. Mobility data usually exist strong repeated patterns and spatiotemporal correlation [26, 27]. Extracting these underlying patterns from high-dimensional tensors is a core step in mobility modelling, and tensor factorization is a powerful tool to achieve this goal. The idea of tensor factorization is to approximate the original tensor by the product of several much smaller matrices/tensors leveraging the low-rank property of the data. Tensor factorization can be considered as a high-order version of matrix factorization or singular value decomposition (SVD). Typical tensor factorization methods include CANDECOMP/PARAFAC (CP) decomposition [28] and tucker decomposition [29] (see [25] for a good introduction of tensor factorization). This section reviews representative works on mobility pattern mining using tensor factorization.

In a two-dimensional case, matrix factorization has been used in understanding mobility patterns, e.g., [30, 31]. In a higher-dimensional setting, an early work [32] organized the taxi trips of Beijing into a three-order (origin × destination × time) tensor and developed a regularized non-negative tucker decomposition method for pattern discovery. Because the factorization of a tensor is often not unique, point of interest (POI) data were used in [32] as auxiliary information to regularize the decomposition. The results showed meaningful spatial clusters and temporal structures of taxi trips. Based on different types of mobility data and regularization methods, research similar to [32] includes [33, 34]. Besides using a single matrix or tensor, [35] used a coupled decomposition to detect transit pattern, where an origin×destination matrix and an origin×transfer×destination tensor share common factor loadings. The coupled approach increases the robustness of the decomposition.

The above research solves tensor factorization by minimizing the reconstruction error. On the other hand, tensor factorization can be considered from a probabilistic perspective and solved by statistical inferences [36]. For example [37] use Latent Dirichlet Allocation (LDA) to decompose the Bike Sharing System trips of Paris organized into a three-order tensor (origin × destination × time) and [38] applied a probabilistic tucker decomposition to understand mobility patterns of smart card transactions in Singapore. A recent work [39] integrated probabilistic tensor factorization in a Bayesian supervised learning model to predict the service choice of ride-sourcing users. The tensor factorization links the dependent variable (i.e., the type of service) with a large set of predictor variables through interpretable hidden patterns, identifying critical variables for the prediction. Besides analyzing collective mobility patterns, tensor factorization can also be used for individual-specific analysis. Similar to topic models and LDA [40], we can assume a group of people share the same
mobility patterns (in the form of factor matrices) while each person is characterized by a specific distribution over mobility patterns. Under this structure, [41, 42] applied an LDA model to extract latent topics and individual topic distributions from smart card data. The model of [41] can be used to infer unknown destinations in smart card transactions given the origin and trip time. A similar two-dimensional LDA model was applied on a license plate recognition data set for individual travel behavior pattern discovery and anomaly detection [43].

Other than mobility pattern mining, tensor factorization has a wide range of applications in more general spatiotemporal modelling tasks. Such as traffic data imputation [44, 45, 46, 47], anomaly detection [43, 48, 49], and traffic prediction [50, 51, 52, 53].

5 Conclusion

This article has detailed a set of machine learning and data mining methods to exploit mobility data. These data are rich in both spatial and temporal aspects and present some challenges for their exploitation. We have introduced three different frameworks:

- the first one, based on structural time series decomposition, is relevant when a long history of data is available and when the objective is to analyze human mobility behaviors on different time scales;
- the second one aimed at predicting travel demand in the short, medium, or long term; in this case, the issue is to develop machine learning models in a supervised framework for prediction tasks, taking into account the specificities of the domain, i.e., the particularities of the transport network, the irregular temporal sampling, or the links between the temporal series to be predicted;
- the third framework is tensor learning and factorization, which aims to extract underlying patterns from high-dimensional tensors.

The session will provide a range of current research in this area by highlighting the potential of machine learning and data mining methods to improve the movement of people in increasingly crowded cities with sustainable development criteria.

References


