Extracting dynamic mobility patterns by Hankel dynamic modes decomposition

Xudong Wang, Lijun Sun*

Department of Civil Engineering, McGill University, Montreal, Canada
xudong.wang2@mail.mcgill.ca, lijun.sun@mcgill.ca
* Corresponding author

Extended abstract submitted for presentation at the 11th Triennial Symposium on Transportation Analysis conference (TRISTAN XI)
June 19-25, 2022, Mauritius Island

April 11, 2022

Keywords: Mobility pattern, dynamic mode decomposition, smart card data, anomaly detection

1 INTRODUCTION

The metro system is becoming an essential component of metropolises, with 182 cities in 56 countries operating metros, carrying an average a total of 168 million passengers per day in 2017 (UITP, 2018). The corresponding high-dimensional spatiotemporal mobility data allow us to understand and solve crucial topics in traffic operation and management, such as mobility pattern discovery, boarding/alighting passenger flow prediction, anomaly detection, to mention but a few. However, it is challenging to capture the critical features and model the high-dimensional data due to the complex dependence and interaction among space and time.

As the spatiotemporal mobility data can be naturally summarized into a multidimensional array, i.e., matrix/tensor, many existing studies apply dimension reduction techniques, such as principal component analysis (PCA) and matrix/tensor decomposition, to reveal the mobility patterns (Sun & Axhausen, 2016). The principle of these methods is projecting the high-dimensional data into a low-dimensional space. They use observations instead of the dynamics of the traffic system to analyze the latent patterns, which cannot capture the evolution of the mobility pattern. To fill the gap, we introduce an analytical framework to study the high-dimensional human mobility data from a dynamic perspective. We apply the dynamic mode decomposition (DMD) with Hankel structure on metro boarding passenger flow to uncover the dynamics of mobility patterns and detect abnormal passenger flow by predicting future passenger flow.

The DMD is a data-driven technique that can simultaneously extract dynamically spatial structures called dynamic modes with corresponding temporal evolution from observations (Schmid, 2010). Though the DMD is applied in many domains, e.g., fluid mechanics, video processing, and epidemiology, only a few studies have been published in the transportation area nowadays (Avila & Mezić, 2020, Lehmberg *et al.*, 2021). The DMD fails to formula the dynamics of the traffic system because the spatial dimension is usually much less than the temporal dimension, which cannot fully capture the system dynamics over the whole period. Therefore, we use delay embedding (Hankelization) to enlarge the spatial dimension of the traffic data and obtain meaningful dynamic modes. We apply hierarchical clustering to group the dynamic modes at different timestamps to demonstrate the dynamics of mobility patterns. We also conduct a long-term prediction experiment using extracted dynamic modes to evaluate the effectiveness of the dynamic modes.

2 METHODOLOGY

We denote the traffic variable (e.g., metro passenger flow) collected from N locations/stations at timestamp t as $\boldsymbol{x}_t \in \mathbb{R}^N,\ t = 1, \cdots, T$ and the whole data set as $\boldsymbol{X} = [\boldsymbol{x}_1, \cdots, \boldsymbol{x}_T] \in \mathbb{R}^{N\times T}$. We assume the traffic system is a locally linear dynamical system so that $\boldsymbol{x}_{t+1} \approx \boldsymbol{A}\boldsymbol{x}_t$, where $\boldsymbol{A} \in \mathbb{R}^{N\times N}$ is a dynamic matrix. The DMD aims to compute the leading eigendecomposition of the best-fit linear operator $\boldsymbol{A}$ for all the traffic data $\boldsymbol{X}$. However, the traffic data have rank mismatch problem due to $N \ll T$ in practice; therefore, we construct a Hankel matrix $\boldsymbol{H} \in \mathbb{R}^{NK\times(N-K+1)}$ to enlarge the data dimension. The Hankelization is a useful data augmentation technique that recursively augment the data by repeating portions of them (Wang *et al.*, 2021). Given the delay embedding length K, we can obtain $\boldsymbol{H}$ from $\boldsymbol{X}$ by Hankelization operation $\mathcal{H}$:

$$\mathcal{H}(\boldsymbol{X}) := \boldsymbol{H}_{(k-1)N+1:kN,\,:} = \boldsymbol{X}_{:,\ t:N-K+t}, \quad \textit{for } k = 1, \ldots, K,\ t = 1, \ldots, T. \tag{1}$$

Similarly, we can obtain $\boldsymbol{h}_{t+1} \approx \boldsymbol{A}_H\boldsymbol{h}_t$, where $\boldsymbol{h}_t \in \mathbb{R}^{NK}$ is the t column of $\boldsymbol{H}$ and $\boldsymbol{A}_H \in \mathbb{R}^{NK\times NK}$ is the dynamic matrix for $\boldsymbol{H}$ (Brunton *et al.*, 2016). So the problem turns to be find the leading eigenvectors and eigenvalues of $\boldsymbol{A}_H$ relating the data $\boldsymbol{H}_2 \approx \boldsymbol{A}_H\boldsymbol{H}_1$ (Tu *et al.*, 2013):

$$\boldsymbol{A}_H = \boldsymbol{H}_2\boldsymbol{H}_1^+, \tag{2}$$

where $\boldsymbol{H}_1 := [\boldsymbol{h}_1, \cdots, \boldsymbol{h}_{N-K}] \in \mathbb{R}^{NK\times(N-K)}$ and $\boldsymbol{H}_2 := [\boldsymbol{h}_2, \cdots, \boldsymbol{h}_{N-K+1}] \in \mathbb{R}^{NK\times(N-K)}$ and $\boldsymbol{H}_1^+$ is the Moore–Penrose inverse of $\boldsymbol{H}_1$.

The eigenvectors named dynamic modes reflect the spatially coherent, and the corresponding eigenvalues determine the time dynamics, e.g., oscillation frequency and decay/growth rate, of these modes. This is the backbone of the DMD. However, it is intractable to analyze $\boldsymbol{A}_H$ directly from Eq. (2) when the dimension is high. Instead, DMD discover the leading eigenvalues and eigenvectors of $\boldsymbol{A}_H$ from a rank-reduced matrix $\tilde{\boldsymbol{A}}_H \in \mathbb{R}^{r\times r}$ by taking truncated SVD of $\boldsymbol{H}_1$ so that $\boldsymbol{H}_1 \approx \boldsymbol{U}_r\boldsymbol{\Sigma}_r\boldsymbol{V}_r^T$ (Tu *et al.*, 2013):

$$\tilde{\boldsymbol{A}}_H = \boldsymbol{U}_r^T\boldsymbol{A}_H\boldsymbol{U}_r = \boldsymbol{U}_r^T\boldsymbol{H}_2\boldsymbol{V}_r\boldsymbol{\Sigma}_r^{-1}, \tag{3}$$

where $\boldsymbol{U} \in \mathbb{R}^{NK\times r}$, $\boldsymbol{\Sigma} \in \mathbb{R}^{r\times r}$, $\boldsymbol{V} \in \mathbb{R}^{(N-K)\times r}$, and r denotes the rank of $\tilde{\boldsymbol{A}}_H$. We use cumulative eigenvalue percentage (CEP) defined as $\sum_{i=1}^{r}\lambda_i/\sum_{i=1}^{\min(NK,N-K)}\lambda_i$ to determine the rank r: the first r eigenvalues are selected when the CEP reaches the threshold δ. It has been proved that $\boldsymbol{A}_H$ and $\tilde{\boldsymbol{A}}_H$ have the same nonzero leading eigenvalues, so we can obtain the eigenvectors $\boldsymbol{\Phi}_H \in \mathbb{C}^{NK\times r}$ and eigenvalues $\boldsymbol{\lambda} \in \mathbb{C}^r$ of $\boldsymbol{A}_H$ by computing the eigenvalue decomposition of $\tilde{\boldsymbol{A}}_H$:

$$\begin{aligned} \tilde{\boldsymbol{A}}_H\boldsymbol{W} &= \boldsymbol{W}\operatorname{diag}(\boldsymbol{\lambda}), \\ \boldsymbol{\Phi}_H &= \boldsymbol{H}_2\boldsymbol{V}_r\boldsymbol{\Sigma}_r^{-1}\boldsymbol{W}, \end{aligned} \tag{4}$$

where columns of $\boldsymbol{W}$ are eigenvectors of $\tilde{\boldsymbol{A}}_H$ (Kutz *et al.*, 2016). We reshape the dynamic modes $\boldsymbol{\Phi}$ to a tensor of size $N \times r \times K$ and apply hierarchical clustering on frontal slices to discover the mobility pattern under different timestamp k.

Given a time period length T_ξ, we can reconstruct the Hankel traffic matrix $\boldsymbol{H}$ and predict the future values by the dynamic modes and eigenvalues:

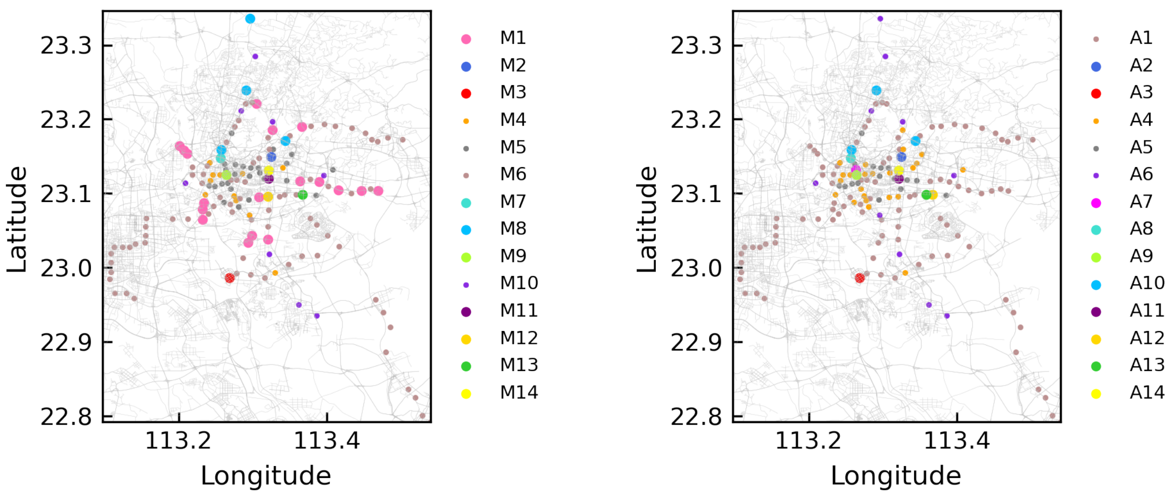
$$\boldsymbol{H}_{\text{DMD}} = \boldsymbol{\Phi}\operatorname{diag}(\boldsymbol{b})\boldsymbol{\Psi}, \tag{5}$$

where $\boldsymbol{b}$ is the mode amplitudes computed as $\boldsymbol{b} = \boldsymbol{\Phi}^+\boldsymbol{h}_1$ and Ψ is a Vandermonde matrix of eigenvalues denoted dynamic evolution, i.e., $\boldsymbol{\Psi}_{i,t} = \lambda_i^{t-1},\ i = 1, \ldots, r,\ t = 1, \ldots, T_\xi$. It has to be noted that the most eigenvalues will decay to zero due to $|\lambda_i| < 1$, which fails to predict long-term future values. To address the problem, we set the norm of some eigenvalues to be 1 when $|\lambda_i| \approx 1$. To the end, we use inverse Hankelization operator $\mathcal{H}^{-1}$ (Wang *et al.*, 2021) to transform the Hankel matrix $\boldsymbol{H}_{\text{DMD}}$ back to $\boldsymbol{X}_{\text{DMD}}$.

3 RESULTS

We apply Guangzhou metro boarding passenger flow collected by a fare collection system (passengers tap in/tap out at fare gantries) to evaluate our model. The data set includes 159 stations and 3 weeks with 15-min resolution from 6:00 to 0:00, i.e., $\boldsymbol{X} \in \mathbb{R}^{159\times(72\times21)}$. We use the first two weeks' data as training data to discover the mobility patterns and the third week to evaluate the prediction performance. We set the Hankel delay embedding length K as 504 (one week) and the rank threshold δ as 0.9 to determine the number of dynamic modes.

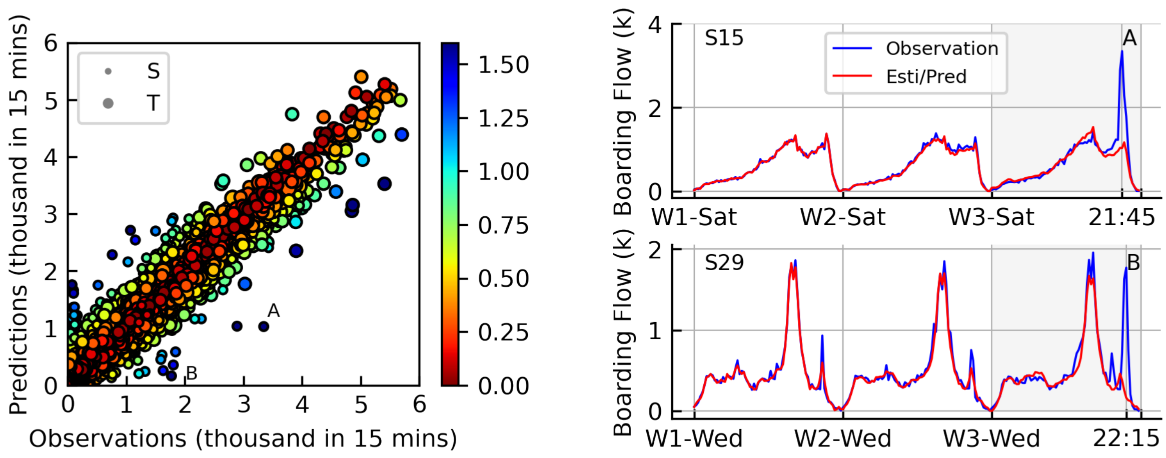
As the dynamic modes describe the spatial coherent, we identify the mobility patterns by hierarchical clustering. Figure 1a and 1b shows the pattern clustering results at 8:00 and 18:00 on Monday, respectively. We can observe from the figure: (1) most stations exhibit spatial correlation regardless of time, such as neighbor stations are clustered as a group (brown cluster and gray cluster); (2) some stations are clustered by area function, e.g., all the stations in the blue cluster are transport hubs; (3) some stations have dynamic spatial correlation; for example, stations in pink group are near the residential area in the morning, carrying passengers to work in the morning, which merges to brown group in the afternoon.



(a) *Mobility pattern clustering at 8:00.* (b) *Mobility pattern clustering at 18:00.*

Figure 1 – *Spatial mobility pattern clustering result under two peak hours. Each dot represents a station and the same color denotes same group.*

The comparison between boarding flow predictions and observations (ground-truth) is shown in Figure 2a. Each dot in (a) represents one pair of observation and its corresponding prediction in every 15 minutes at all the stations. The size of dots represents the station as a single station (S) or transfer station (T), the color of dots denotes the absolute error. It can be seen that most data pairs are predicted accurately (red color) except for several points (blue color). It demonstrates that the extracted dynamic modes can reproduce the passenger flow. To figure out the reason for inaccurate prediction, we select two points with significant prediction errors (Point A and Point B) from 2a. Point A occurs at 21:45 Saturday at Station 15, and Point B at 22:15 Wednesday at Station 29. We plot the predictions (gray background) of these two stations in Figure 2b. The same days of the week in the training data (white background) are shown in the figure as the passenger flow exhibits strong periodicity. It can be seen that unexpected large boarding flow causes the prediction error in both scenarios. In other words, Point A and Point B can be regarded as anomalies detected by the model.



(a) *Observations and predictions for the third week using the extracted leading dynamic modes.*

(b) *The prediction of Point A and Point B denoting in (a).*

Figure 2 – *The prediction and anomaly detection results.*

4 CONCLUSION

In this paper, we apply the dynamic mode decomposition with Hankel structure on metro boarding passenger flow to reveal the dynamics of mobility patterns. The Hankelization of mobility data guarantees that the DMD successfully works on the metro system by enlarging the spatial dimension and obtaining the dynamic modes simultaneously. Unlike other dimension reduction methods, the DMD uncovers the dynamical characteristics of the system; therefore, we can exploit the underlying mobility pattern by separately grouping the dynamic modes of stations at any timestamp. The dynamic correlation between stations can be used as prior information to model the spatiotemporal data in future research. Another aspect where the framework shines is that it can achieve long-term prediction and anomaly detection and deal with data heterogeneity.

References

Avila, AM, & Mezić, I. 2020. Data-driven analysis and forecasting of highway traffic dynamics. *Nature communications*, **11**(1), 1–16.

Brunton, Bingni W, Johnson, Lise A, Ojemann, Jeffrey G, & Kutz, J Nathan. 2016. Extracting spatial–temporal coherent patterns in large-scale neural recordings using dynamic mode decomposition. *Journal of neuroscience methods*, **258**, 1–15.

Kutz, J Nathan, Brunton, Steven L, Brunton, Bingni W, & Proctor, Joshua L. 2016. *Dynamic mode decomposition: data-driven modeling of complex systems*. SIAM.

Lehmberg, Daniel, Dietrich, Felix, & Köster, Gerta. 2021. Modeling Melburnians—Using the Koopman operator to gain insight into crowd dynamics. *Transportation Research Part C: Emerging Technologies*, **133**, 103437.

Schmid, Peter J. 2010. Dynamic mode decomposition of numerical and experimental data. *Journal of fluid mechanics*, **656**, 5–28.

Sun, Lijun, & Axhausen, Kay W. 2016. Understanding urban mobility patterns with a probabilistic tensor factorization framework. *Transportation Research Part B: Methodological*, **91**, 511–524.

Tu, Jonathan H, Rowley, Clarence W, Luchtenburg, Dirk M, Brunton, Steven L, & Kutz, J Nathan. 2013. On dynamic mode decomposition: Theory and applications. *arXiv preprint arXiv:1312.0041*.

UITP. 2018. *World metro figures 2018–statistic brief*. https://cms.uitp.org/wp/wp-content/uploads/2020/06/Statistics-Brief-World-metro-figures-2018V3_WEB.pdf.

Wang, Xudong, Wu, Yuankai, Zhuang, Dingyi, & Sun, Lijun. 2021. Low-Rank Hankel Tensor Completion for Traffic Speed Estimation. *arXiv preprint arXiv:2105.11335*.