Understanding large-scale traffic flow using model-based and data-driven dimension reduction: with COVID-19 and Olympic-Paralympic case study

Toru Seo[a],*[1]

[a] *Department of Civil and Environmental Engineering, Tokyo Institute of Technology*

Abstract: Traffic flow in urban road networks is an extremely large, high-dimensional phenomenon, and it is generally difficult to gain an intuitive understanding of it. For example, there is no established way to directly answer questions such as how desirable a state is, how similar one state is to another. This makes it difficult to evaluate and manage urban traffic. This study proposes use of model-based and data-driven dimension reduction methods (i.e., Macroscopic Fundamental Diagram and Uniform Manifold Approximation and Projection) to obtain intuitive understanding of traffic. Our analysis of actual highway and arterial data in Tokyo in 2019 and 2021 (with COVID-19 and Olympic-Paralympic) show that regular and irregular patterns can be easily understood using these methods.

Key-words: Traffic flow, Macroscopic Fundamental Diagram, UMAP, COVID-19, Olympic-Paralympic

Introduction

Urban-scale traffic is a very large and complex phenomenon. Typically, vehicular traffic flows occur over large areas of an entire city with resolution on the order of 10 to 100 meters, and occur throughout the year with resolution on the order of minutes to hours. It is difficult to intuitively recognize and interpret the overall state of such a large and detailed phenomenon. This difficulty in interpretation poses challenging issues in traffic science and engineering. This is not only a problem for humans, but also for statistical and machine learning methods in the form of a curse of dimensionality.

In general, dynamic network traffic state can be quantified as a vector of flows, densities, and speeds for all links (or enough locations) at a given time. However, this vector is difficult to interpret due to the high-dimensionality of the vector and non-linearity of traffic phenomena. For example, the Euclidean distance between two vectors does not have essential meaning on traffic because it does not distinguish the direction of traffic. It would be desirable to have a method to compute the intrinsic similarity and distance between traffic states.

To overcome the difficulties caused by this high dimensionality, the Macroscopic Fundamental Diagram (MFD) is used by the transportation research community (Geroliminis and Daganzo, 2007). It represents the state of a road network in terms of two-dimensional state quantities: the network average flow and density. Under idealized conditions, this allows correct and concise recognition of the state of the traffic system and enables simple and efficient traffic control through area inflow control, for example.

However, there are several problems with the low-dimensional representation of traffic phenomena by MFD. First, MFD can represent essentially different traffic phenomena as the same state quantity. For example, a morning peak and an evening peak, during which the direction of traffic flow are completely opposite, may have the same average network flow and density and be indistinguishable on MFD. Furthermore, in order to obtain well-defined MFD, the network should satisfy some homogeneity conditions.

MFD can be interpreted as a model-based dimension reduction method. On the other hand, data-driven dimension reduction methods may provide a more appropriate (in terms of data representation) low-dimensional representation of traffic phenomena. With similar motivations, data-driven clustering methods have been widely applied to traffic phenomena to understand their features (e.g., Wang et al,

([1]) Corresponding author (seo.t.aa@m.titech.ac.jp)

2012; Saeedmanesh and Geroliminis, 2017; Zhang et al., 2022). However, to the author's knowledge, continuous dimension reduction methods (e.g., principal component analysis) have not been widely applied to network traffic flows to understand their nature. Because traffic phenomena are inherently continuous, continuous dimension reduction methods may have substantial advantages.

The purpose of this study is to propose a novel approach to intuitively recognize and interpret large-scale traffic phenomena. The approach consists of a model-based dimension reduction method (MFD) and a data-driven method called Uniform Manifold Approximation and Projection (UMAP) (McInnes et al., 2018), a relatively new method. The approach is then applied to actual traffic data obtained in Tokyo in 2019 (ordinally state) and 2021 (with COVID-19 and Olympic-Paralympic Games) to derive some insights on ordinally traffic as well as very special traffic and comparison between these two.

Methodology and Data

UMAP is a non-linear dimension reduction method that computes low-dimensional representation of given data based on manifold learning. It is known that the method is capable of representing global structure of the data while preserving the local structure well. Its computational cost is very low.

The summary of the method is as follows. Let $\boldsymbol{x}_t$ be a vector of traffic data collected on time t, X be a set of $\boldsymbol{x}_t$ for all t. Then, 2-dimensional vector can be obtained as $\boldsymbol{u}_t = UMAP_X(\boldsymbol{x}_t)$ using UMAP (represented as function $UMAP_X$ for dataset X) or $\boldsymbol{m}_t = MFD(\boldsymbol{x}_t)$ using MFD (represented as function MFD which computes the network averages). Finally, the distributions of $\boldsymbol{u}_t$ and $\boldsymbol{m}_t$ are investigated.

We apply MFD and UMAP to traffic data obtained in Tokyo. Specifically, we use the following datasets:

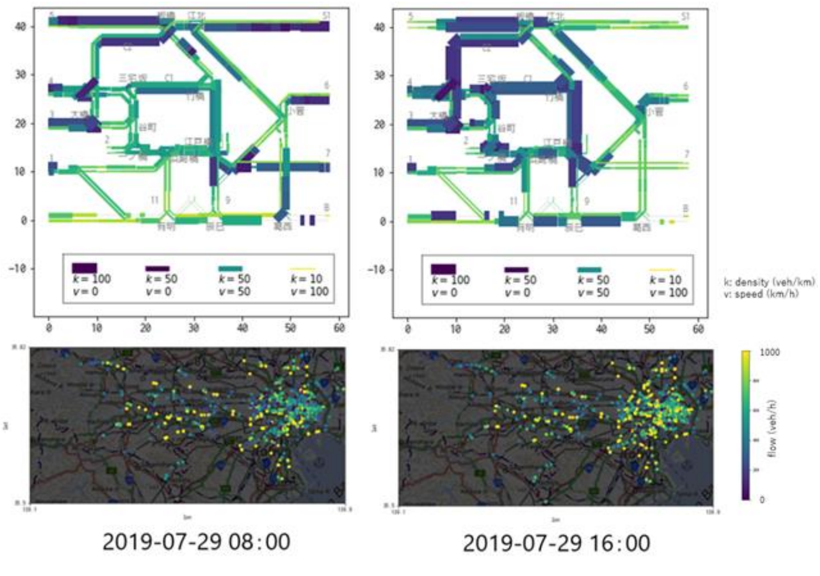
— 5-min flow and density measured by traffic detectors at Tokyo Metropolitan Expressway. The total number of available detectors is 1425.

— 5-min flow measured by traffic detectors at major arterial roads in Tokyo. The total number of available detectors is 626.

— Data collection periods: from July to September in 2019 and 2021. The total number of 5-min time slots (i.e., the size of X) is about 25000. The data in 2021 is thoroughly affected by COVID-19. Olympic Game was conducted from 2021-07-23 to 08-08 and Paralympic Game was conducted from 2021-08-24 to 09-05. The traffic regulation for the Games is summarized at https://www.2020games.metro.tokyo.lg.jp/special/eng/traffic/, but actual impact is not clarified.

See **Figure 1** for the illustration of the traffic data. Unfortunately, speed data is not available in arterial road. Thus, MFD cannot be computed on it; therefore, for arterial road, only UMAP was computed.

Figure 1. Network traffic states on highway (up) and major arterial (bottom) on morning and evening peaks.

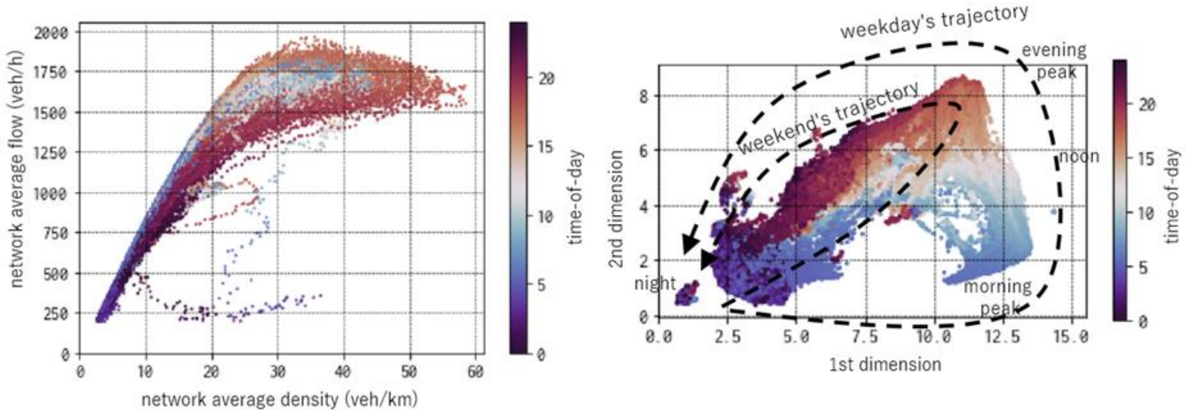


Sourse: Created by the author using graphical data provided by Geospatial Information Authority of Japan, Mr. Takara Sakai, and Mr. Yutaro Ishikawa

Results

MFD and UMAP on the highway are shown in **Figure 2**. Each dot represents $\boldsymbol{m}_t$ or $\boldsymbol{c}_t$ on specific t, and the color represents the time-of-day of t. The shape of MFD can be considered as standard for highways. The shape of UMAP can be expressed as "distorted donut with a gap at 6 o'clock position". In fact, this donut-like shape has very clear physical interpretation. As seen from the color, $\boldsymbol{c}_t$ on a certain (week)day moves on a counterclockwise trajectory on this diagram during the day. Furthermore, on weekend or holiday, a day's trajectory does not pass the lower-right region in the diagram. This means that a qualitative difference exists between traffic patterns in weekdays and weekends, and UMAP is useful to quantify it. This difference was not observed in MFD.

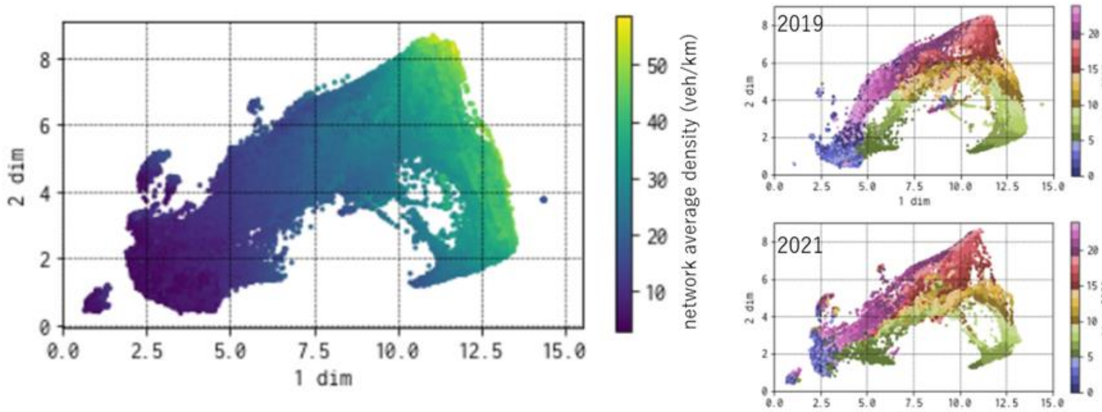
Figure 2. 2-dimentional representations of network traffic state using MFD (left) and UMAP with annotations (right) on the highway.



Source: own elaborations

Figure 3 shows further details on UMAP. It was confirmed that MFD's network average density is well correlated to UMAP-representation; it means that the network average density is very useful variable to represent the network state, from both of model-based and data-driven perspectives. On the contrary, no clear relation was observed between average flow and UMAP. By comparing $\boldsymbol{c}_t$ from 2019 and 2021, slight differences were observed before and during COVID-19.

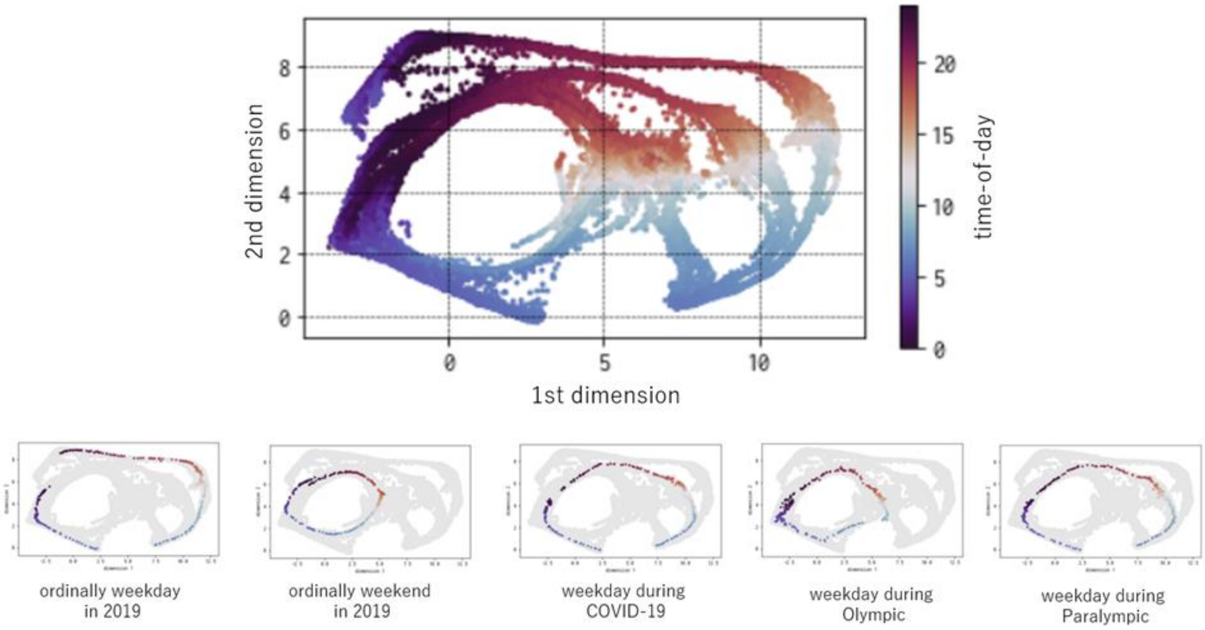
Figure 3. 2-dimentional representations of network traffic state using UMAP and MFD's network average density (left) and UMAP in 2019 and 2021 (right) on the highway.



Source: own elaborations

UMAP on the arterial roads is shown in **Figure 4**. You can see a remarkable pattern consists of multiple donut-like shapes. Similar to the highway result, each donut-like shape corresponds to a trajectory of within-day traffic for a certain period: ordinally weekday in 2019, ordinally weekend in 2019, day during COVID-19, and day during the Games. See the bottom of **Figure 4** for trajectories of typical days. These results suggests that UMAP can recognize qualitative differences of traffic patterns in an intuitive way.

Figure 4. 2-dimentional representations of network traffic state using UMAP on the arterial roads (top: all results, bottom: trajectories of typical days).



Source: own elaborations

Conclusions

The key insights from this study are as follows. First, the UMAP results showed similar donut-like shapes at the two different locations. We could hypothesize that this is a universal phenomenon in transportation. Second, UMAP is capable of distinguishing and quantifying the qualitative differences in traffic pattens, such as weekday and weekend, before and during COVID-19 pandemic, and with and without the Games. Especially, the results on arterial roads would be remarkable because of its intuitiveness and simpleness; to the authors' knowledge, this kind of results have not been reported in the literature. Third, MFD's network average density is meaningful to represent the network state, from both of model-based and data-driven perspectives. Finally, traffic patterns on highway were not significantly affected by COVID-19 or the Games. On the other hand, ones on arterial were significantly affected by the both of COVID-19 and the Games differently. The future works includes in-depth investigation of the results and application to traffic prediction and control.

Acknowledgements

The data was freely provided by Tokyo Metropolitan Government ("offering of several datasets on mobility and transportation in Tokyo 2020 Games"), JARTIC, Takara Sakai, and Yutaro Ishikawa. Funding was provided by JSPS KAKENHI Grant-in-Aid 20H02267.

References

Geroliminis, N. and Daganzo, C. F.: Macroscopic modelling of traffic in cities, *Transportation Research Board 86th Annual Meeting*, 2007.

McInnes, L., Healy, J., and Melville, J.: UMAP: Uniform manifold approximation and projection for dimension reduction, *arXiv preprint arXiv:1802.03426*, 2018.

Wang, P., Hunter, T., Bayen, A. M., Schechtner, K., González, M. C.: Understanding road usage patterns in urban areas. *Scientific Reports*, Vol. 2, No. 1001, 2012.

Saeedmanesh, M., Geroliminis, N.: Dynamic clustering and propagation of congestion in heterogeneously congested urban traffic networks. *Transportation Research Part B: Methodological* Vol. 105, pp. 193–211, 2017.

Zhang, P., Ma,W., Qian, S.: Cluster analysis of day-to-day traffic data in networks. *Transportation Research Part C: Emerging Technologies*, Vol. 144, 103882, 2022.