Generalization of localized linear regression time series model to freeway networks

Semin Kwak and Nikolas Geroliminis

École polytechnique fédérale de Lausanne, 1015 Lausanne, Switzerland,
{semin.kwak, nikolas.geroliminis}@epfl.ch

1 Introduction

Traffic prediction is one of the most important features to support successful Intelligent Transportation System (ITS). Accurate traffic prediction not only can enable for operators to establish dynamic control strategies but also for a traveller to make decisions about his or her trip by having useful information such as travel time so that which route would take the least travel time under dynamic conditions.

Literature review in the topic is vast, but only a few papers are cited due to space limitation. Yildirimoglu and Geroliminis [1] grouped the historical data set to three clusters (weekdays, Friday, and weekend), and build the stochastic congestion map for each cluster. Based on the stochastic map, the authors matched the real-time measurement to the most similar congestion map with a threshold and predict traffic state. Chandra and Al-Deek [2] predicted traffic by applying vector autoregressive (VAR) model. Although the authors studied VAR model with traffic data and the results show great performances, they conclude that VAR model is highly dependent on the stationary condition of traffic data, which means it is possible that the assumption is not realistic. Also the authors said the performance degradation of prediction would be occur since the rush hour behavior would be different from the off-peak, while time series models use global parameters including VAR model. Nanthawichit *et al.* [3] predicted traffic state (velocity field) by using the macroscopic fundamental diagram (MFD) model. There are also several works based on neural network framework [4]–[6].

Reference [7] introduced a localized linear regression time series model to predict traffic state and experienced travel times. The authors showed that the model captures nonlinear traffic propagation well for a single freeway, not a network, regardless of the traffic states and their dynamics, such as congestion or free flow state. They showed the model only for a single freeway route so, in this paper, we introduce a way for generalizing the model to freeway networks which creates significant computational and modeling challenges.

One issue with extending to network cases comes from the massive number of data points. For example, the freeway network in the district 7 of California has around 900 detectors as shown in Fig. 1. The higher data dimension forces the larger training data set for training huge transition matrices in the model. Another thing to keep in mind is to take an advantage of spatial correlation

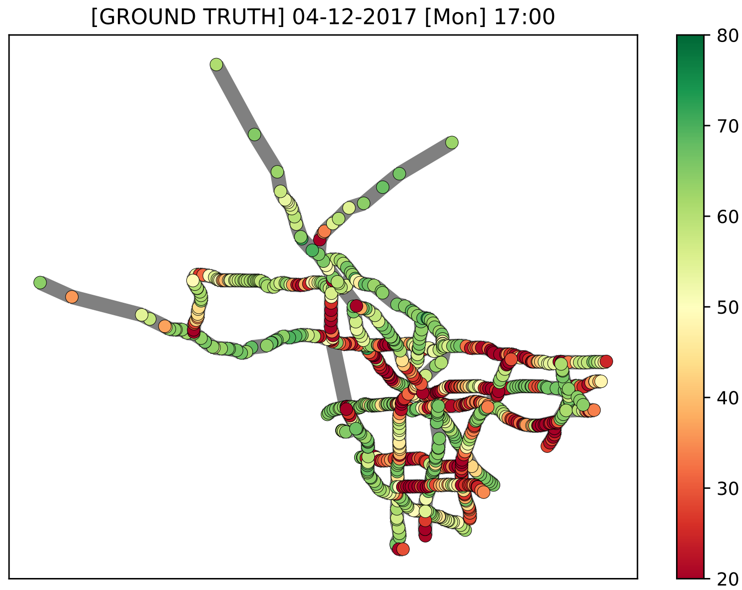

Fig. 1. The traffic state (speed) of December 4^{th} (Monday), 2017 of district 7 in California.

correlations among links in the network. For example, links in the center area of Fig. 1 are more interacting to each other than those in boundaries.

Therefore, in this work, we introduce a well known data dimensional reduction method in graph theory that exploits the Laplacian matrix which is used for transforming a traffic state from vertex domain to spectral domain. This transformation allows us to reduce the dimension of data by considering the structure of a freeway network. We show that it is possible to take the full advantage of the localized linear transformation on the graph spectral domain so that prediction is successfully performed under freeway networks.

2 Localized linear regression time series model

In the work of [7], the authors introduced a localized linear regression time series model.

$$\mathbf{x}_{t+1}^d = \mathbf{H}_{t+1,t}\mathbf{x}_t^d + \mathbf{n} \tag{1}$$

The authors built this model based on three important hypothesis: First, current state ($\mathbf{x}_t^d$) is linearly transformed to the next state ($\mathbf{x}_{t+1}^d$) with the transition matrix ($\mathbf{H}_{t+1,t}$). Second hypothesis is that the transformation is localized in time domain which means that the transformation is only valid between t and $t+1$. Lastly, the authors assumed that the transformation is independent of days, regardless of different traffic demands.

By assuming each element of the noise vector follows Gaussian distribution independently and identically, the authors trained the transition matrix minimizing the square error of noise matrix (collection of noise vectors for different days), and the final solution is:

$$\hat{\mathbf{H}}_{t+1,t} = \mathbf{X}_{t+1}\mathbf{X}_t^T(\mathbf{X}_t\mathbf{X}_t^T)^{-1} \tag{2}$$

where the matrix $\mathbf{X}_t$ represents the collection of state vectors which corresponds the time t in the historical data set.

The authors also designed a linear predictor by minimizing mean square error between predicted state and the real state as follows:

$$\mathbf{x}_t^d(h) = \mathbf{H}_{t+h,t}\mathbf{x}_t^d, \tag{3}$$

where $\mathbf{x}_t^d(h)$ represents the h-time step ahead predictor at time t and $\mathbf{H}_{t+h,t} = \prod_{i=1}^{h} \mathbf{H}_{t+h-i+1,t+h-i}$.

3 Data dimensional reduction for graph signal

The main issue with extending from a freeway to a network comes from the expansion of data dimension. In equation (2), the computational cost of the matrix multiplication depends on the size of the state matrix. Also, the matrix $\mathbf{X}_t\mathbf{X}_t^T$ would be rank deficient when the data dimension is higher than the number of training data.

Data dimension reduction methods have a potential to alleviate such issues. Here, we introduced the unnormalized Graph Laplacian which is defined as $\mathbf{L} := \mathbf{D} - \mathbf{W}$, where the degree matrix $\mathbf{D}$ is a diagonal matrix whose ith diagonal element is equal to sum of the weights of all the edges incident to vertex i. Therefore, for any graph signal $\mathbf{x}$ which is defined in a vertex domain, it satisfies:

$$(\mathbf{L}\mathbf{x})(i) = \sum_{j \in N_i} \mathbf{W}(i,j)\left[\mathbf{x}(i) - \mathbf{x}(j)\right] \tag{4}$$

where the neighborhood N_i is the set of vertices connected to vertex i by an edge, the ith element of vector $\mathbf{x}$ is represented as $\mathbf{x}(i)$, and the ith row and jth column entry of matrix $\mathbf{W}$ is represented as $\mathbf{W}(i,j)$. Because the Laplacian matrix is symmetric by the definition, it has a complete set of orthonormal eigenvectors ($\mathbf{U}$) and corresponding eigenvalues ($\mathbf{\Lambda}$):

$$\mathbf{L} = \mathbf{U}\mathbf{\Lambda}\mathbf{U}^T \tag{5}$$

where (i,i)-th element of the diagonal matrix $\mathbf{\Lambda}$ represents ith eigenvalue in ascending order, i.e., $\mathbf{\Lambda}(1,1) \leq \mathbf{\Lambda}(2,2) \leq \cdots \leq \mathbf{\Lambda}(N,N)$ and ith column of the matrix $\mathbf{U}$ represents the eigenvector which corresponds to the eigenvalue $\mathbf{\Lambda}(i,i)$. If a graph is connected, then $\mathbf{\Lambda}(1,1) = 0$.

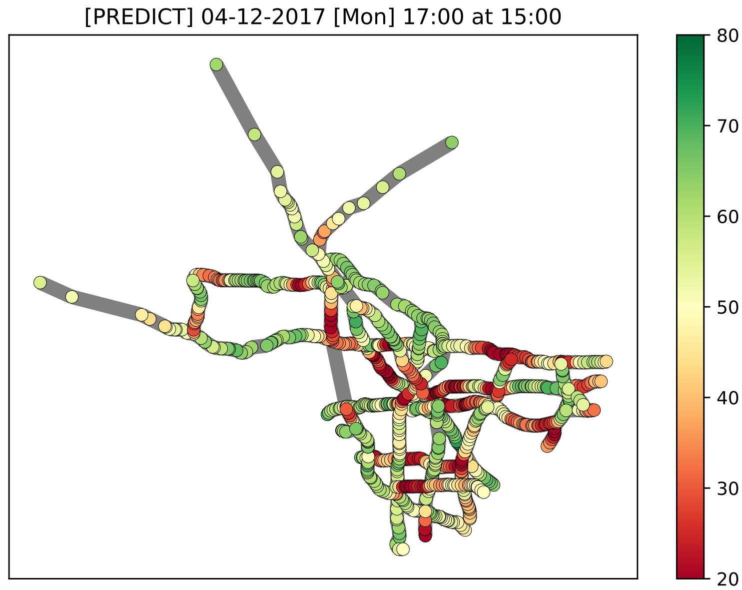

Fig. 2. The prediction of the traffic state (speed) of 17:00 at 15:00, December $4th$ (Monday) 2017, district 7 of California.

We transform the original graph signal (in our case, traffic state vector $\mathbf{x}$) to the row space of $\mathbf{U}$

$$\hat{\mathbf{x}} = \mathbf{U}^T \mathbf{x} \tag{6}$$

Interestingly, the transformed vector $\hat{\mathbf{x}}$ contains the information of "spectrum" of the state vector $\mathbf{x}$ with the information of the graph structure which is defined by the weight matrix $\mathbf{W}$. For example, the first element of the vector $\hat{\mathbf{x}}\,(1)$ contains dc (direct current) information of the signal $\mathbf{x}$, but the last element of the vector $\hat{\mathbf{x}}\,(N)$ contains the information of much higher frequency of the signal $\mathbf{x}$. For more information, refer the paper [8].

For the data compression, we chose subset of eigenvectors rather than the full set of those

$$\tilde{\mathbf{x}} = \tilde{\mathbf{U}}^T \mathbf{x} \tag{7}$$

where $\tilde{\mathbf{U}} = \mathbf{U}\,(:, 1 : k)$.

Therefore, we applied localized linear regression time series model with the transformed and compressed state vector $\tilde{\mathbf{x}}$ instead of $\mathbf{x}$ in equation (1). After prediction by equation (3) with the compressed spectral signal $\tilde{\mathbf{x}}$, the predicted spectral signal $\tilde{\mathbf{x}}_t\,(h)$ is reconstructed as follows:

$$\mathbf{x}_t\,(h) = \tilde{\mathbf{U}} \tilde{\mathbf{x}}_t\,(h) \tag{8}$$

4 Numerical results

We used 300 basis (eigenvectors), so the compression rate is $300/900 = 1/3$ and trained the transition matrices $\mathbf{H}_{t+1,t}$ with the traffic data (speed) for the freeway network in district 7 of California from 2015-01-01 to 2017-11-30. Figure 2 shows one example of prediction result at 15:00 of 2017-12-04 for 2 hours later (17:00).

Comparing the result to the ground truth (Fig. 1), it is shown that the prediction based on the localized linear regression time series model by compressing data performs well. Since we choose 300 eigenvectors which correspond to 300 smallest eigenvalues, we can see that the prediction result seems smoother that the ground truth. It means that the traffic signal is naturally denoised (similar to low pass filter) during prediction process. Also, because the dc component of the traffic signal (average of traffic values of a network at a certain time) is always dominant to other components, it is unnecessary to execute the post process which bounds the traffic signal to have a proper value, e.g., non negative speed.

5 Conclusive remarks

In this paper, we generalized localized linear regression time series model to network case. We introduced graph spectral transformation to reduce data dimension to make it possible to train the transition matrix. In the full paper, more test data will be examined to evaluate the performance of the traffic predictor and be compared with the state-of-the-art techniques. For the future work, further generalization of the method to urban networks is a research priority that contains additional challenges.

References

[1] M. Yildirimoglu and N. Geroliminis, "Experienced travel time prediction for congested freeways," *Transportation Research Part B: Methodological*, vol. 53, pp. 45–63, Jul. 2013.

[2] S. R. Chandra and H. Al-Deek, "Predictions of freeway traffic speeds and volumes using vector autoregressive models," *Journal of Intelligent Transportation Systems*, vol. 13, no. 2, pp. 53–72, 2009.

[3] C. Nanthawichit, T. Nakatsuji, and H. Suzuki, "Application of probe-vehicle data for real-time traffic-state estimation and short-term travel-time prediction on a freeway," *Transportation Research Record: Journal of the Transportation Research Board*, no. 1855, pp. 49–59, 2003.

[4] C.-H. Wei and Y. Lee, "Development of freeway travel time forecasting models by integrating different sources of traffic data," *IEEE Transactions on vehicular technology*, vol. 56, no. 6, pp. 3682–3694, 2007.

[5] P. V. Palacharla and P. C. Nelson, "Application of fuzzy logic and neural networks for dynamic travel time estimation," *International Transactions in Operational Research*, vol. 6, no. 1, pp. 145–160, 1999.
[6] N. G. Polson and V. O. Sokolov, "Deep learning for short-term traffic flow prediction," *Transportation Research Part C: Emerging Technologies*, vol. 79, pp. 1–17, 2017.
[7] S. Kwak and N. Geroliminis, "Experienced travel time forecasting method for freeways with a localized linear regression time series model," *IEEE Transactions on Intelligent Transportation Systems*, submitted.
[8] D. I. Shuman, S. K. Narang, P. Frossard, A. Ortega, and P. Vandergheynst, "The emerging field of signal processing on graphs: Extending high-dimensional data analysis to networks and other irregular domains," *arXiv preprint arXiv:1211.0053*, 2012.