

A Brief Overview of Machine Learning Methods for Short-term Traffic Forecasting and Future Directions

Yaguang Li, Cyrus Shahabi

Department of Computer Science, University of Southern California
{yaguang, shahabi}@usc.edu

Abstract

Short-term traffic forecasting is a vital part of intelligent transportation systems. Recently, the combination of unprecedented data availability and the rapid development of machine learning techniques have brought on immense advancement in this field. In this paper, we aim to provide a brief overview of machine learning approaches for short-term traffic forecasting to facilitate research in related fields. We first introduce traffic forecasting and the challenges, and then introduce different approaches for modeling the temporal and/or spatial dependencies. Finally, we discuss several important directions for the future research.

1 Introduction

Traffic forecasting is the core component of intelligent transportation systems (ITS). The goal of traffic forecasting is to estimate future traffic conditions of a transportation network based on historical observations. Based on the forecasting horizon, traffic forecasting can be categorized as short-term forecasting and long-term forecasting. In this paper, we will focus on short-term traffic forecasting, whose forecasting horizon is usually less than or equal to one hour. Short-term traffic forecasting is important for various applications, including route planning [17], traffic control [10], car dispatching [1], etc.

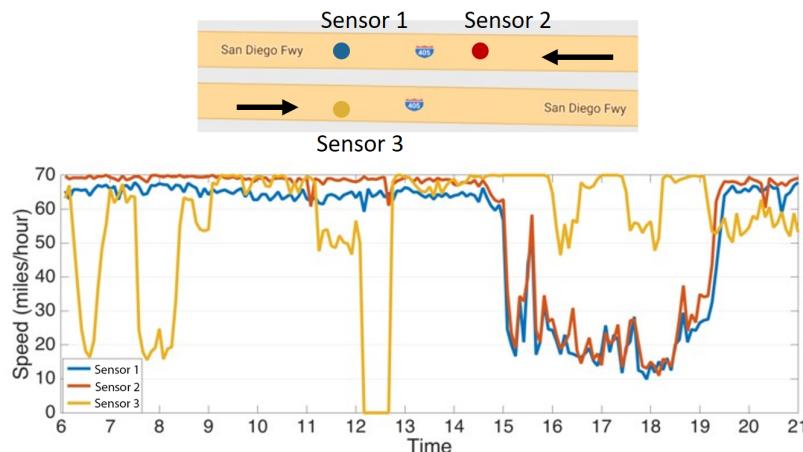


Figure 1: Complex spatial dependency among different traffic time series. Reprinted from [19] with permission.¹

This problem is challenging mainly due to the **complex spatial and temporal dependencies** [19]. On the one hand, traffic time series demonstrate strong **temporal dynamics**. Recurring incidents such as **rush hours** or **accidents** can result in formation of non-stationary time series, rendering forecast challenging. On the other hand, sensors on the road network contain complex yet unique **spatial correlations**. Figure 1 illustrates an example of spatial dependencies among different traffic time series. Suppose there are traffic sensors on three roads, i.e., sensor 1, 2 and 3. The traffic time series of sensor 1 and 2 are correlated, while those of sensor 1 and sensor 3 are not. Though road 1 and road 3 are close in the Euclidean space, they demonstrate very different behaviors.

Traffic forecasting has been studied in various communities ranging from transportation system [29, 31], through economics [28, 9], and to data mining [21, 20], and its methods mainly fall into two categories: **knowledge-driven approach** and **data-driven approach**. In transportation and operational research, knowledge-driven methods usually try to computationally **model the transportation network** through queuing theory and simulating driver behaviors in traffic [8, 4]. With the availability of increasing amount of traffic data, **data-driven machine learning approaches for traffic forecasting have received considerable attention**. In this paper, we will give a brief overview of different data-driven machine learning approaches for short-term traffic forecasting and describe potentially future directions in this field.

2 Overview of Short-term Traffic Prediction Approaches

In this section, we give a brief overview of different traffic forecasting approaches. These approaches are categorized into two types based on **whether they model the spatial correlation among different traffic time series**.

2.1 Traffic Forecasting without Modeling Spatial Dependency

Traffic forecasting can be modeled as **a time series regression problem** and thus various time series analysis approaches have been applied to this problem.

Historical Average models the traffic flow as a seasonal process, and uses the weighted average of previous seasons as the prediction. For example, suppose the season is 1 week, then the prediction for this Wednesday is the averaged traffic speeds from last four Wednesdays. **As the historical average method does not depend on short-term data, its performance is invariant to the small increases in the forecasting horizon**. Auto-regressive integrated moving average (**ARIMA**) is a popular model for time series analysis and has been successfully applied to traffic forecasting [16]. ARIMA consists of three parts: 1) the Auto-regressive (**AR**) part indicates that the evolving variable of interest can be approximated using a linear combination of its own historical values, 2) the Moving average (**MA**) part is used to model the residual from the AR part using a weighted combination of random noises at various previous time steps, and 3) the Integrate (**I**) part models the difference between adjacent values rather than raw values. In [34], the authors use **Seasonal ARIMA** to capture the periodicity in the traffic flow, while in [26] **ARIMA is augmented with historical average to better model the rush hour traffic behavior**.

Other popular time series methods for traffic forecasting include K-nearest Neighbor (**KNN**) [42, 3], Support Vector Regression (**SVR**) [30], particle filter, **Hidden Markov Model** [27], **Gaussian Process** [36], etc. However, these time series models usually rely on the stationary assumption, which is often **violated by the real-world traffic data**.

To model the non-linear temporal dependency, **neural network based approaches** have also been applied to traffic forecasting. In [22, 12], the authors propose to use **stacked denoising encoder and deep belief networks** to model the temporal behavior. In [24, 40, 15], the authors model the temporal dependency using Recurrent

¹For this and all subsequent reprinted figures, the original authors retain their copyrights, and permission was obtained from the corresponding author.

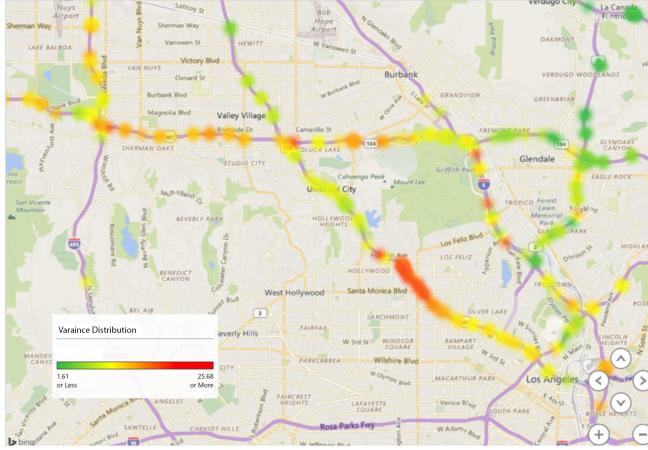


Figure 2: Distribution of standard deviation from the historical average. The hardness of traffic prediction varies across different locations.

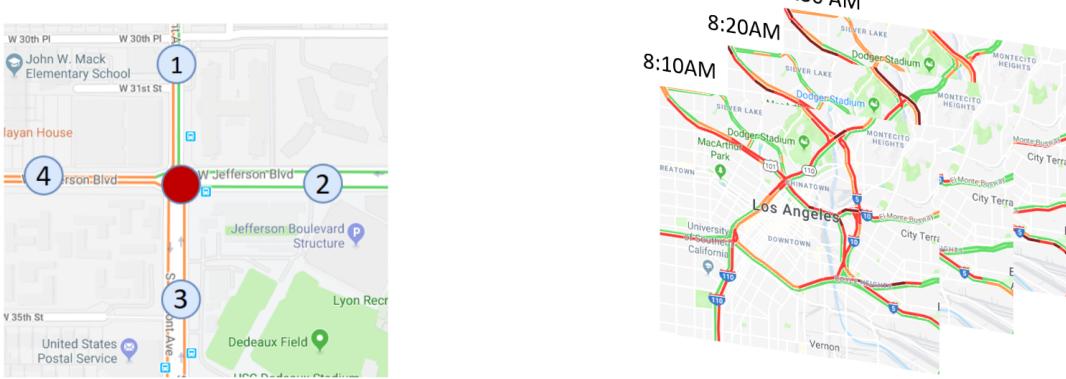
Neural Networks (**RNN**), which is a type of neural network with self-connection, and is able to perform non-linear auto-regression. However, the majority of the above-mentioned approaches model each traffic time series separately, failing to capture the spatial dependency among them.

2.2 Traffic prediction with Modeling Spatial Dependency

To capture the spatial dependency among traffic time series, researchers have extended existing approaches to process multivariate time series. The resulted models include **Vector Auto-regressive** [11], **Vector ARIMA** [13], **Spatiotemporal ARIMA** [25], **Spatiotemporal HMM** [14, 37]. In [6], the authors further propose to first group similar sensors and then perform multi-task learning on each group. An alternative way to model the relationship among different time series is the latent space model which **first transforms the raw traffic time series into the latent space and then learns the spatiotemporal dependency**. In [39], the authors propose a **temporal regularized matrix factorization based approach** which performs vector auto-regression in the latent space. While in [7], the authors model the **road network as a graph**, and propose to **learn the attributes of vertices** in latent spaces which captures both topological and temporal properties.

However, existing machine learning models either impose strong stationary assumptions on the data (e.g., **auto-regressive model**) or fail to **account for highly non-linear temporal dependency** (e.g., latent space model [39, 7]). Deep learning models deliver new promise for time series forecasting problem. To capture the spatial dependency of the traffic, recent studies [35, 23, 41] propose to model the transportation network as an image and use Convolutional Neural Networks (**CNN**) to extract spatial features. One drawback of these methods is that they **ignore the topology of the underlying transportation network**, e.g., in Figure 1, two roads in different directions of a highway, though close in Euclidean distance, can have significantly different traffic pattern because of the network topology.

To resolve this issue, Li et al. [19] model the underlying road network as a directed weighted graph and propose diffusion convolutional recurrent neural network (**DCRNN**) which systematically captures the topological dependency using diffusion convolution. Diffusion convolution is a new form of **convolutional operation defined on the graph based on the diffusion nature of traffic**. Later, in [38], the authors speed up this model by **replacing RNN with CNN** to model the temporal dependency. In [5], the authors introduce **DeepTransport**, which models the spatial dependency by **explicitly collecting upstream and downstream neighborhood roads** for each individual road and conduct convolution on these neighborhoods. The methods discussed above are not exhaustive, and more related work and details can be found in a recent survey paper [32] and references therein.



(a) Which neighbor road has the largest effects?

(b) Which historical observation is more relevant?

Figure 3: Interpretable traffic forecasting.

3 Discussion and Future Directions

In this section, we present several future directions for traffic forecasting.

Traffic prediction in extreme cases While traffic patterns in normal conditions are easy to predict, a more interesting question in traffic forecasting is to forecast traffic for extreme conditions, which include both peak hours and post-accident traffic forecasting. In [40], the authors propose to learn a representation of accident features with auto-encoder and then combine with recurrent neural network for post-accident traffic forecasting. While improved performance is observed, the proposed model fails to consider the correlation among different sensors and the results can still be improved further.

Fuse traffic prediction with other applications Many important applications in transportation are strongly related to traffic prediction. One example is travel time estimation (ETA) [18]. Currently, traffic prediction and travel time estimation are usually performed independently. It is desirable to have a model that jointly models these two problems and achieves improved results for either task.

Long-term temporal dependency modeling Very long-term temporal dependency usually exists in traffic data, e.g., the current traffic situation can be strongly correlated with a day, a week or even several months ago. Currently, the most popular approaches to model the non-linear temporal dependency is recurrent neural networks (RNN). However, due to the sequential nature of RNN, it is hard to model very long term dependencies [33]. Besides, RNN is not efficient to train as it is hard to parallelize. Thus efficient approaches that are able to capture long-term non-linear temporal dependencies are much needed.

Evaluation metric design Popular metrics to evaluate traffic forecasting include Mean Absolute Error (MAE), Root Mean Squared Error (RMSE) and Mean Absolute Percentage Error (MAPE) which are calculated by averaging across all sensors. These metrics put an equal importance on all sensors and time slots. However, we argue not all sensors and time slots are equally informative w.r.t. evaluating the performance. Figure 2 shows each location's standard deviation from its historical average. Generally, the larger the standard deviation is the harder to predict the traffic at that location. Arguably, locations and times with higher error, e.g., busy intersections during peak-hours, are more important to predict. Alternatively, predicting the average speed on

all freeways from midnight to 5am is not very difficult. Thus, it might be beneficial to have a metric that give more rewards to predicting traffic at tougher locations and times.

Interpretable traffic prediction Many machine learning models are used for traffic prediction. Though, good performance is achieved, the prediction made by the model are usually not interpretable. As shown in Figure 3, it is desirable to identify which spatial and temporal components affect the model’s prediction. Besides, rather than a single prediction, it is more informative to predict a distribution, e.g., the mean and the variance of the Gaussian distribution, which would help decision making as well as other related applications, e.g., travel time estimation [2].

Acknowledgments

This research has been funded in part by NSF grant CNS-1461963, Caltrans-65A0533, the USC Integrated Media Systems Center (IMSC), and the USC METRANS Transportation Center. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of any of the sponsors such as NSF.

References

- [1] M. Asghari, D. Deng, C. Shahabi, U. Demiryurek, and Y. Li. Price-aware real-time ride-sharing at scale: an auction-based approach. In *Proceedings of the 24th ACM SIGSPATIAL International Conference on Advances in Geographic Information Systems*, page 3. ACM, 2016.
- [2] M. Asghari, T. Emrich, U. Demiryurek, and C. Shahabi. Probabilistic estimation of link travel times in dynamic road networks. In *Proceedings of the 23rd SIGSPATIAL International Conference on Advances in Geographic Information Systems*, page 47. ACM, 2015.
- [3] P. Cai, Y. Wang, G. Lu, P. Chen, C. Ding, and J. Sun. A spatiotemporal correlative k-nearest neighbor model for short-term traffic multistep forecasting. *Transportation Research Part C: Emerging Technologies*, 62:21–34, 2016.
- [4] E. Cascetta. *Transportation systems engineering: theory and methods*, volume 49. Springer Science & Business Media, 2013.
- [5] X. Cheng, R. Zhang, J. Zhou, and W. Xu. Deeptransport: Learning spatial-temporal dependency for traffic condition forecasting. *arXiv preprint arXiv:1709.09585*, 2017.
- [6] D. Deng, C. Shahabi, U. Demiryurek, and L. Zhu. Situation aware multi-task learning for traffic prediction. In *Data Mining (ICDM), 2017 IEEE International Conference on*, pages 81–90. IEEE, 2017.
- [7] D. Deng, C. Shahabi, U. Demiryurek, L. Zhu, R. Yu, and Y. Liu. Latent space model for road networks to predict time-varying traffic. In *SIGKDD*, pages 1525–1534, 2016.
- [8] D. R. Drew. Traffic flow theory and control. Technical report, 1968.
- [9] G. Duranton and M. A. Turner. The fundamental law of road congestion: Evidence from us cities. *The American Economic Review*, pages 2616–2652, 2011.
- [10] J. F. Gilmore and N. Abe. Neural network models for traffic control and congestion prediction. *Journal of Intelligent Transportation Systems*, 2(3):231–252, 1995.

- [11] J. D. Hamilton. *Time series analysis*, volume 2. Princeton university press Princeton, 1994.
- [12] W. Huang, G. Song, H. Hong, and K. Xie. Deep architecture for traffic flow prediction: deep belief networks with multitask learning. *ITS, IEEE Transactions on*, 15(5):2191–2201, 2014.
- [13] Y. Kamarianakis and P. Prastacos. Forecasting traffic flow conditions in an urban network: Comparison of multivariate and univariate approaches. *Transportation Research Record: Journal of the Transportation Research Board*, (1857):74–84, 2003.
- [14] J. Kwon and K. Murphy. Modeling freeway traffic with coupled hmms. Technical report, Technical report, Univ. California, Berkeley, 2000.
- [15] N. Laptev, J. Yosinski, L. E. Li, and S. Smyl. Time-series extreme event forecasting with neural networks at Uber. In *Int. Conf. on Machine Learning Time Series Workshop*, 2017.
- [16] S. Lee and D. Fambro. Application of subset autoregressive integrated moving average model for short-term freeway traffic volume forecasting. *Transportation Research Record: Journal of the Transportation Research Board*, (1678):179–188, 1999.
- [17] Y. Li, D. Deng, U. Demiryurek, C. Shahabi, and S. Ravada. Towards fast and accurate solutions to vehicle routing in a large-scale and dynamic environment. In *International Symposium on Spatial and Temporal Databases*, pages 119–136. Springer, 2015.
- [18] Y. Li, K. Fu, Z. Wang, C. Shahabi, J. Ye, and Y. Liu. Multi-task representation learning for travel time estimation. In *International Conference on Knowledge Discovery and Data Mining (KDD)*, 2018.
- [19] Y. Li, R. Yu, C. Shahabi, and Y. Liu. Diffusion convolutional recurrent neural network: Data-driven traffic forecasting. In *International Conference on Learning Representations (ICLR)*, 2018.
- [20] M. Lippi, M. Bertini, and P. Frasconi. Short-term traffic flow forecasting: An experimental comparison of time-series analysis and supervised learning. *ITS, IEEE Transactions on*, 14(2):871–882, 2013.
- [21] W. Liu, Y. Zheng, S. Chawla, J. Yuan, and X. Xing. Discovering spatio-temporal causal interactions in traffic data streams. In *SIGKDD*, pages 1010–1018. ACM, 2011.
- [22] Y. Lv, Y. Duan, W. Kang, Z. Li, and F.-Y. Wang. Traffic flow prediction with big data: A deep learning approach. *ITS, IEEE Transactions on*, 16(2):865–873, 2015.
- [23] X. Ma, Z. Dai, Z. He, J. Ma, Y. Wang, and Y. Wang. Learning traffic as images: a deep convolutional neural network for large-scale transportation network speed prediction. *Sensors*, 17(4):818, 2017.
- [24] X. Ma, Z. Tao, Y. Wang, H. Yu, and Y. Wang. Long short-term memory neural network for traffic speed prediction using remote microwave sensor data. *Transportation Research Part C: Emerging Technologies*, 54:187–197, 2015.
- [25] W. Min and L. Wynter. Real-time road traffic prediction with spatio-temporal correlations. *Transportation Research Part C: Emerging Technologies*, 19(4):606–616, 2011.
- [26] B. Pan, U. Demiryurek, and C. Shahabi. Utilizing real-world transportation data for accurate traffic prediction. In *Data Mining (ICDM), 2012 IEEE 12th International Conference on*, pages 595–604. IEEE, 2012.
- [27] Y. Qi and S. Ishak. A hidden markov model for short term prediction of traffic conditions on freeways. *Transportation Research Part C: Emerging Technologies*, 43:95–111, 2014.

- [28] K. A. Small and E. T. Verhoef. *The economics of urban transportation*. Routledge, 2007.
- [29] B. L. Smith and M. J. Demetsky. Traffic flow forecasting: comparison of modeling approaches. *Journal of transportation engineering*, 123(4):261–266, 1997.
- [30] H. Su, L. Zhang, and S. Yu. Short-term traffic flow prediction based on incremental support vector regression. In *Natural Computation, 2007. ICNC 2007. Third International Conference on*, volume 1, pages 640–645. IEEE, 2007.
- [31] E. I. Vlahogianni, J. C. Golias, and M. G. Karlaftis. Short-term traffic forecasting: Overview of objectives and methods. *Transport reviews*, 24(5):533–557, 2004.
- [32] E. I. Vlahogianni, M. G. Karlaftis, and J. C. Golias. Short-term traffic forecasting: Where we are and where were going. *Transportation Research Part C: Emerging Technologies*, 43:3–19, 2014.
- [33] E. Vorontsov, C. Trabelsi, S. Kadoury, and C. Pal. On orthogonality and learning recurrent networks with long term dependencies. In *ICML*, 2017.
- [34] B. M. Williams and L. A. Hoel. Modeling and forecasting vehicular traffic flow as a seasonal arima process: Theoretical basis and empirical results. *Journal of transportation engineering*, 129(6):664–672, 2003.
- [35] Y. Wu and H. Tan. Short-term traffic flow forecasting with spatial-temporal correlation in a hybrid deep learning framework. *arXiv preprint arXiv:1612.01022*, 2016.
- [36] Y. Xie, K. Zhao, Y. Sun, and D. Chen. Gaussian processes for short-term traffic volume forecasting. *Transportation Research Record: Journal of the Transportation Research Board*, (2165):69–78, 2010.
- [37] J. Xu, D. Deng, U. Demiryurek, C. Shahabi, and M. Van Der Schaar. Mining the situation: Spatiotemporal traffic prediction with big data. *IEEE Journal of Selected Topics in Signal Processing*, 9(4):702–715, 2015.
- [38] B. Yu, H. Yin, and Z. Zhu. Spatio-temporal graph convolutional neural network: A deep learning framework for traffic forecasting. *IJCAI*, 2018.
- [39] H.-F. Yu, N. Rao, and I. S. Dhillon. Temporal regularized matrix factorization for high-dimensional time series prediction. In *Advances in Neural Information Processing Systems*, pages 847–855, 2016.
- [40] R. Yu, Y. Li, C. Shahabi, U. Demiryurek, and Y. Liu. Deep learning: A generic approach for extreme condition traffic forecasting. In *SIAM International Conference on Data Mining (SDM)*, 2017.
- [41] J. Zhang, Y. Zheng, and D. Qi. Deep spatio-temporal residual networks for citywide crowd flows prediction. In *AAAI*, pages 1655–1661, 2017.
- [42] L. Zhang, Q. Liu, W. Yang, N. Wei, and D. Dong. An improved k-nearest neighbor model for short-term traffic flow prediction. *Procedia-Social and Behavioral Sciences*, 96:653–662, 2013.