Transfer Learning for Traffic Speed Prediction

Paper: 20-01604

Bo Wang[1], Inhi Kim[1], Hai L. Vu[1]

[1] Monash University, Department of Civil Engineering, Australia. Bo.Wang1@monash.edu

INTRODUCTION

Neural Networks have achieved great performance in traffic prediction, but they need abundant data for supervised learning. In practice, the systematic replacement or expansion in traffic-related systems usually needs new sensors, which lack historical data for the model training. In this case, learned knowledge from the longer used sensors may help the forecasting accuracy for newly installed sensors by the application of the fine-tuning technique. This research uses the fine-tuning method for traffic speed prediction, and the average improvement of the control group and experimental group is 7.4% and 39.6%, respectively (based on the MSE).

MODEL STRUCTURE

The proposed model uses multiple traffic variables for modeling. The data shape of the model inputs and outputs are different, which can be denoted as:

$$\text{Input Data Shape} = [N, \quad S, \quad T_{in}, \quad F_{in}]$$

$$\text{Output Data Shape} = [N, \quad S, \quad T_{out}, \quad F_{out}]$$

where N is batch size, S is the number of detectors, T_{in} and T_{out} are input and output time-steps, and F_{in} and F_{out} are the numbers of input and output traffic attributes.

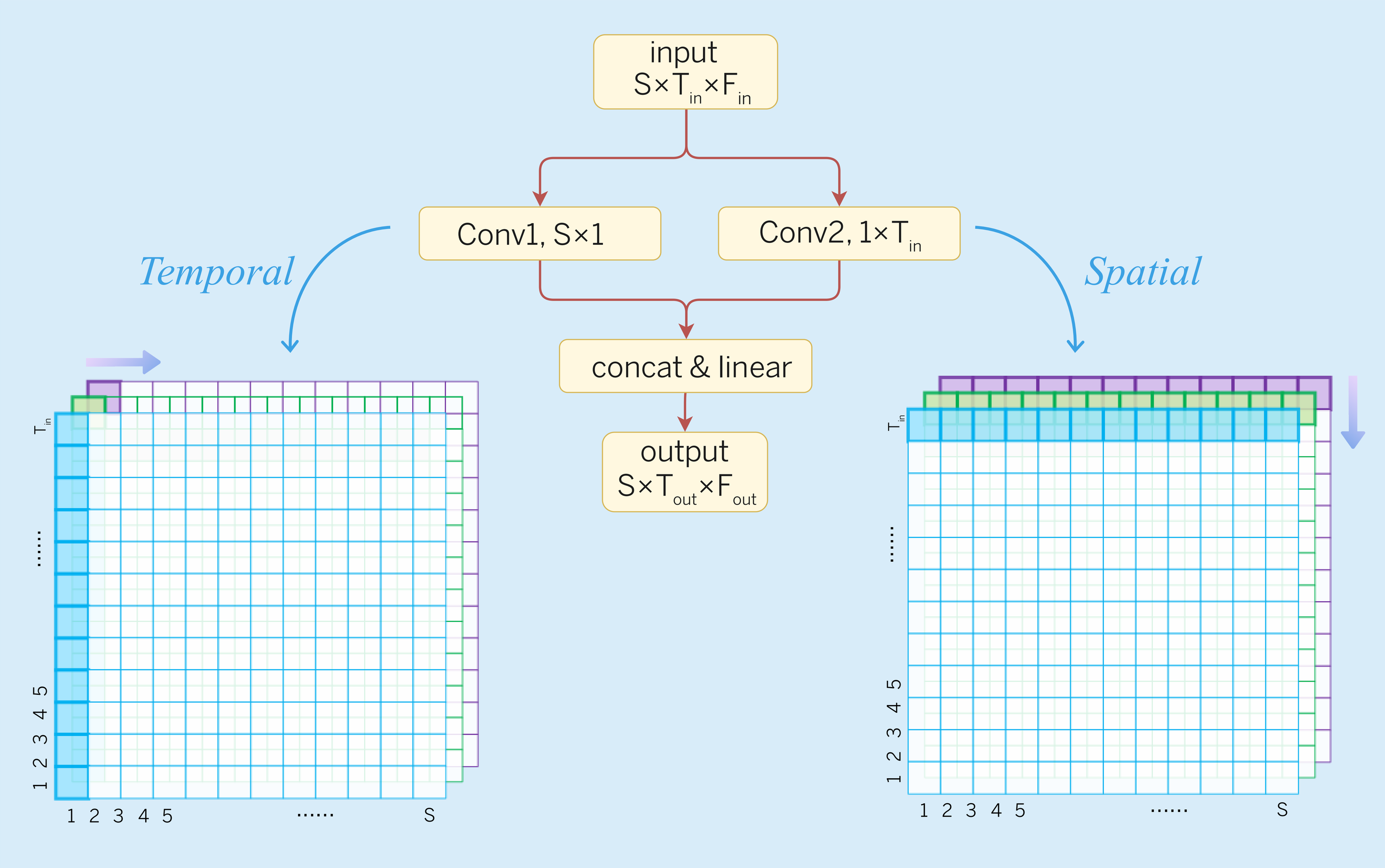

FIGURE 1 Model Structure and Two Different Kernels

Two kernels move through different time-steps and locations, respectively. After each CNN layer, the ReLU function would be performed. Finally, the linear layer concatenates all outputs from the previous two layers, then performs linear regression for correct output size.

DATA DESCRIPTION

The data source in this study is the Caltrans Performance Measurement System (PeMS), which stores multiple traffic variables over 39,000 individual traffic detectors in San Diego, California, US.

FIGURE 2 Detector Locations of two datasets and their input data shape

Two subsets are created, and multiple traffic data of speed, density, and flow are used for forecasting the traffic speed in the next {5,10,15,20} continues mins. Table 1 shows the comprehensive information on these two datasets.

TABLE 1 Dataset Information

Dataset	Modeling	Case Study
Detectors	12	12
Traffic direction	Northbound	Northbound
Postmile	[34.033, 41.922]	[43.191, 51.327]
Time Interval	5 mins	5 mins
Start	1/03/2018 0:00	1/03/2018 0:00
End	31/05/2018 23:55	7/03/2018 23:55
Length	27,648	2,016
Test Set	Last 300 time-steps	Last 300 time-steps
Training Set	Random 90%	Random 90%
Validation Set	Random 10%	Random 10%

WORKFLOW OF EXPERIMENTS

All models appear in the figure share the same structure, but they use different methods to initialize the hidden weights and biases before training. The "Random" method generates trainable parameters from a uniform distribution, while the "Fine-tuning" method (B) initializes them from a trained model (A). For the second method, the learned traffic features (A) are expected to be transferred to a new case (B).

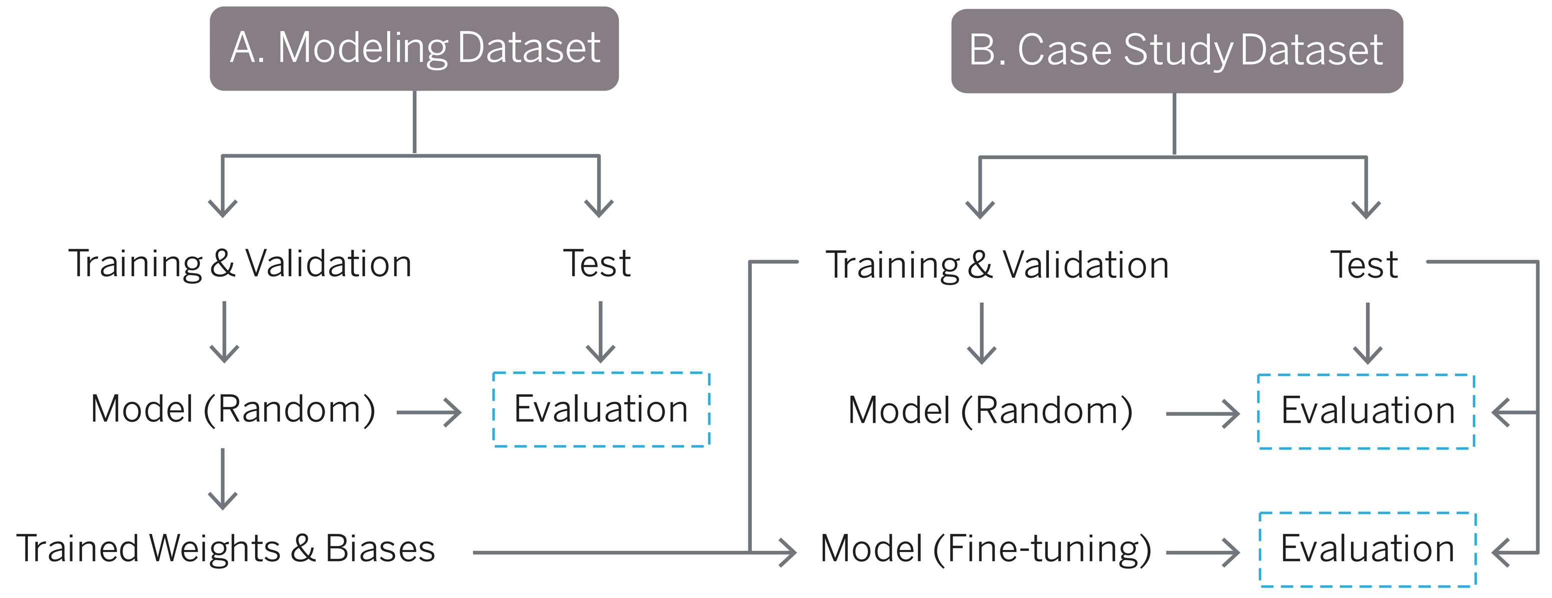

FIGURE 3 The workflow of transfer learning experiments

RESULTS

Table 2 shows the detailed results of the experiments. Each group performs model 3 training and evaluation using two different datasets. The data duration of the Modeling and Case 4 Study models are three months and one week, respectively.

TABLE 2 Forecasting Performance

Dataset	Output	Control Group (CNN+LSTM) Params	MSE	RMSE	MAE	MAPE	Experimental Group (Two Parallel CNN) Params	MSE	RMSE	MAE	MAPE
A. Modeling		Random					Random				
	5 mins	238792	57.118	4.489	1194	0.122	18020	2.862	1.518	367	0.021
	10 mins	463060	57.814	4.562	2450	0.123	32432	4.332	1.849	893	0.026
	15 mins	687328	58.717	4.666	3817	0.125	32432	5.657	2.056	1516	0.03
	20 mins	911596	59.198	4.699	5149	0.126	46844	6.977	2.235	2206	0.032
B. Study Case		Random (7.4% improved)					Random (39.6% improved)				
	5 mins	238792	44.926	4.432	1144	0.091	18020	8.078	2.614	723	0.038
	10 mins	463060	46.879	4.657	2376	0.094	32432	9.625	2.803	1505	0.041
	15 mins	687328	47.351	4.688	3608	0.094	32432	11.452	2.988	2403	0.044
	20 mins	911596	48.637	4.734	4897	0.095	46844	12.13	3.087	3240	0.044
		Fine-tuning					Fine-tuning				
	5 mins	238792	42.183	4.12	1044	0.086	18020	3.721	1.65	434	0.024
	10 mins	463060	42.733	4.179	2148	0.087	32432	5.818	2.132	1124	0.031
	15 mins	687328	44.549	4.381	3433	0.091	32432	7.55	2.243	1722	0.033
	20 mins	911596	44.485	4.336	4487	0.09	46844	8.399	2.355	2452	0.035

These results indicate that:

1. Fewer training data lead to worse performance in both control and experimental groups.
2. The results in the experimental group used fewer parameters and produced better results than the control group, which indicates that the proposed two parallel CNNs outperformed the CNN-LSTM structure
3. The learned experience can be transferred for other locations (from A to B), and the fine-tuning results of the Case Study are improved with limited training data.

DISCUSSION

There are four subplots in Figure4 for four different output durations: {5,10,15,20} mins. In each subplot, the x-axis represents the training epoch, and the y-axis represents the MSE of the corresponding validation set. The fine-tuning weights are adaptive for all output durations, and they help the model converge much faster than when random starting values (see all subplots) are used.

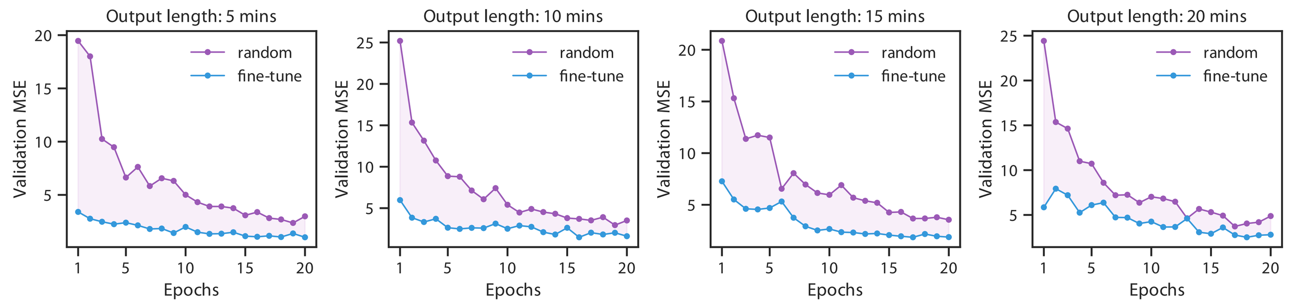

FIGURE 4 Comparison of Model Initialization Methods

Figures 5 (a) and (b) demonstrate the RMSE and real traffic speed distribution of the test set in the Modeling model. The error distribution varies significantly with different hours. We note two issues of RMSE preventing the model from further improvement: 1. The location 39.793 around 10:30 AM; 2. The general fluctuation between 14:00 and 19:00 PM. At the same time, the corresponding speed in subplot (b) also changes largely.

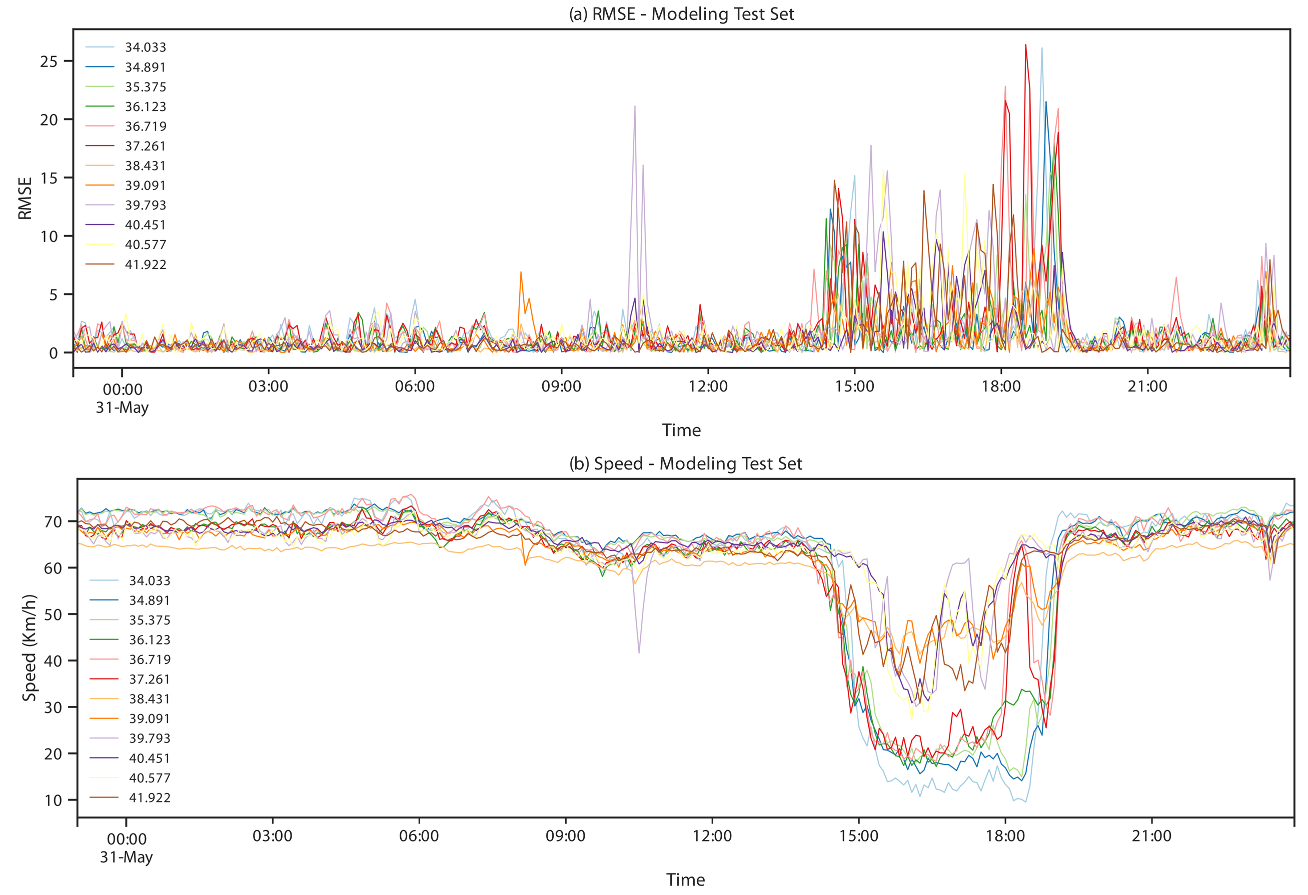

FIGURE 5 The RMSE and speed distribution in the test set

ACKNOWLEDGMENTS

Accessing to the high-performance computing system at Monash (Massive 3 or M3) to process and analyze the data for this study is gratefully acknowledged.