Short-term Traffic Flow Prediction with Deep Neural Networks and Adaptive Transfer Learning

Junyi Li, Fangce Guo, Yibing Wang, Lihui Zhang, Xiaoxiang Na, Simon Hu *Member, IEEE*

Abstract—**A key problem in short-term traffic prediction is the prevailing data missing scenarios across the entire traffic network. To address this challenge, a transfer learning framework is currently used in the literature, which could improve the prediction accuracy on the target link that suffers severe data missing problems by using information from source links with sufficient historical data. However, one of the limitations in these transfer-learning based models is their high dependency on the consistency between datasets and the complex data selection process, which brings heavy computation burden and human efforts. In this paper, we propose an adaptive transfer learning method in short-term traffic flow prediction model to alleviate the complex data selection process. Specifically, a self-adaptive neural network with a novel domain adaptation loss is developed. The domain adaptation loss is able to calculate the distance between the source data and the corresponding target data in each training batch, which can help the network to adaptively filter inconsistent source data and learn target link related information in each training batch. The Maximum Mean Discrepancy (MMD) measurement, which has been fully validated and applied in transfer learning research, is used in combination with the Gaussian kernel to measure the distance between datasets in each training batch. A series of experiments are designed and conducted using 15-minute interval traffic flow data from the Highways England, UK. The results have demonstrated that the proposed adaptive transfer learning method is less affected by the inconsistency between datasets and provides more accurate short-term traffic flow prediction.**

Keywords—**Short-term Traffic Prediction, Deep Neural Networks, Adaptive Transfer Learning**

I. Introduction

Short-term traffic prediction has been recognized as an essential part in the Intelligent Transportation System (ITS), which allows the application of various traffic control strategies to properly assign the vehicles and efficiently employ the existing road network. Existing studies in short-term traffic prediction range from statistical methods like regression model, Kalman Filtering (KF) and ARIMA [1]-[5] to machine learning methods like Support Vector Regression (SVR), k-Nearest Neighbors (kNN) and Neural Networks (NN) [6]-[11]. For example, Kamarianakis & Prastacos [1] compared the traffic flow forecasting performance of four classic statistical models using arterial loop detector data in Athens, Greece. Guo et al. [11] investigated the application of SVR, kNN and Grey System Model (GSM) in short-term traffic prediction and compared prediction accuracy using traffic flow and occupancy data collected from the urban area of London, UK. The Machine Learning(ML) methods have gained a great popularity in solving the short-term traffic prediction problem and proven to be effective in prediction under different traffic demands and boundary conditions. In this paper, we focused on using the Deep Neural Networks (DNN) in traffic prediction as they are not only less complicated to implement, but also potentially able to capture more unstable non-linear traffic patterns hence provide more accurate short-term predictions.

Despite the well-established techniques in short-term traffic prediction, most of existing studies in traffic prediction focused on corridor-level or small network-level problems. For traffic prediction in an entire large network, it is not easy to obtain sufficient traffic data from all sensor sites, resulting in prevailing data missing scenarios and causing extra heavy computational burden. In practice, it is impossible to collect high-quality data in all minor links across a huge network with thousands of links nor calibrate such a huge model with different sub-models on each link. Although some research successfully applied Convolutional Neural Network (CNN) to learn the successive traffic states in an area as images, they failed to consider the problems of data insufficiency in several weeks and months [12]-[14]. Moreover, the training time of such networks can be up to several times than other algorithms like kNN and Recurrent Neural Network (RNN) [12], which is a huge barrier to real-time application and on-line learning in practice.

Another potential solution to address the problems of data insufficiency and heavy computation overhead is to improve the model transferability between sub-models in different links. Consequently, the calibrated transferable model can be applied to a series of links after slightly fine-tuning towards the target link. In our previous work [15], we combined a Long Short-Term Memory (LSTM) network with transfer learning techniques to calibrate a shared model that generalizes intermediate patterns of traffic state variables of a set of similar links. Specifically, in the first step, a source network is created using sufficient data from the source links; in the next step, layer parameters are partly extracted and transferred to create the target network after fine-tuning towards the insufficient target link data. The model sensitivity analysis and field studies showed that although the proposed hybrid method is able to improve the prediction accuracy of the target link

J, Li and F, Guo are with Department of Civil and Environmental Engineering, Imperial College London, London SW7 2AZ, UK. (e-mail: junyi.li18@imperial.ac.uk; fangce.guo@imperial.ac.uk).

*S, Hu is with ZJU-UIUC Institute, Zhejiang University, Zhejiang 314400, China (e-mail: simonhu@zju.edu.cn, phone: +8618867516624).

Y, Wang, L, Zhang and *S, Hu are with Institute of Intelligent Transportation Systems, School of Civil Engineering and Architecture, Zhejiang University, Zhejiang 310058, China (email: wangyibing@zju.edu.cn; lihuizhang@zju.edu.cn; simonhu@zju.edu.cn).

Xiaoxiang Na is with Department of Engineering, University of Cambridge, Trumpington Street, Cambridge, CB2 1PZ, UK (email: xnhn2@eng.cam.ac.uk)

Manuscript 269 submitted to 2020 IEEE Intelligent Transportation Systems Conference (ITSC). Received March 7, 2020.

under severe data missing conditions, it is highly dependent on consistency between the source link dataset and target link dataset. In the prediction of [15], all the candidate source links with sufficient data should be evaluated and selected according to a complex statistical comparison to the existing target link information, which brings about further computation burden and human efforts.

Given the existing works and identified research gaps above, the main objective of this paper is to propose an Adaptive Transfer Learning (ATL) method based on our previous transfer learning framework in [15] to alleviate the rigorous data selection criteria to predict short-term traffic flow. The network training process is adaptive in accordance with the distance between the source data and the corresponding target data in each training batch, thus the massive data selection work can be significantly reduced. With the proposed methodology, a series of field studies are conducted to test the optimal hyperparameters of this model and validate the efficacy of this ATL method.

The rest of the paper is organized as follows: Section II reviews the previous transfer learning framework in [15] and describes the Self-adaptive Neural Network (SNN), the domain adaptation loss and the Maximum Mean Discrepancy (MMD) measurement. Section III presents a case study in which this proposed network is tested and analyzed. Finally, the conclusions and outlooks are discussed in Section IV.

II. METHODOLOGY

A. *Preliminaries*

To make our concepts clear, some notions are firstly defined as follows:

Source dataset: a dataset with sufficient labelled training data collected from the source link.

Source network: a neural network calibrated by the source dataset.

Target dataset: a dataset collected from the target link, which suffers severe data insufficiency problems.

Target network: a neural network transferred from the first n layers of the source network and finely tuned towards the target dataset.

Data consistency: the data consistency between source dataset and target dataset is evaluated via the cross-correlation coefficient, which measures the overall data patterns between datasets. The consistent source dataset has a high cross-correlation coefficient towards the target dataset whereas the inconsistent source dataset can not match the target dataset very well.

B. *Previous transfer learning framework*

The transfer learning framework in our previous work in [15] can be simply demonstrated as follows. Firstly, a set of source links with sufficient historical data is manually selected based on link geographical attributes and data patterns between source links and target link. Secondly, a deep neural network with three LSTM layers and one fully connected layer is empirically selected and trained by using source dataset only. In the next step, the first n layers of the source network are transferred to the target network. Finally, the target network is finely tuned towards the insufficient target dataset via different layer-parameter fine-tuning strategies. An example with 3 transferred layers is shown in Figure 1.

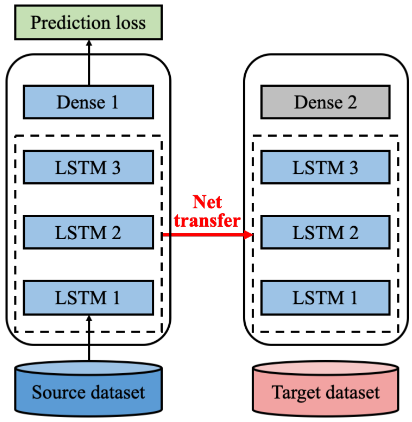

Figure 1. A transfer example with 3 transferred layers

C. *Self-adaptive Neural Network*

As described above, the previous transfer learning framework in Figure 1 is dependent on consistency between datasets that massive manual selection efforts and heavy computation burden are inevitable. To address this challenge, many scholars add an adaptation layer in their proposed neural network to adaptively learn from source data from different feature space and different distribution [16]-[19], which is on the cutting-edge of current transfer learning techniques. The similar notion is applied in this research to achieve the adaptation between the source dataset and the target dataset, and finally enhance the generalization capability of the network.

Many studies have been conducted on exploring the potential domain adaptation method in deep neural networks, and a well-established technique is to re-define the loss function in the training process:

$$l = l_c(D_s, y_s) + \lambda l_A(D_s, D_t) \quad (1)$$

where l is the overall loss of the network; $l_c(D_s, y_s)$ represents the conventional loss in the regression task, which is calculated as the mean squared error (MSE) between the network prediction $f_{NN}(D_s)$ and the ground truth y_s ; $l_A(D_s, D_t)$ represents the loss from adaptation, where the loss evaluation criteria will be discussed in the next section; D_s is the source dataset, D_t is the target dataset; λ is the parameter that weighs the two parts.

The adaptation loss $l_A(D_s, D_t)$ is able to calculate the distance between the source data and the corresponding target data in each training batch. In this case, the data consistency evaluation process is automatically embedded in this network. Thus, the source data with high consistency towards the target data can be adaptively selected and learned. It is worth to note that our Self-adaptive Neural Network is different from current transfer learning methods in image classification area, where most scholars used the domain adaptation methods in the target network to assist the final fine-tuning process. However, the proposed Self-adaptive Neural Network in this work aims to apply domain adaptation methods in the source network to calibrate a shared model that generalizes intermediate patterns of traffic state variables of a set of similar links.

The notion of the proposed Self-adaptive Neural Network in this paper is shown in Figure 2 and Figure 3. Different from our previous transfer learning method in [15], the proposed Self-adaptive Neural Network re-defined the overall loss function as the summation of the conventional prediction loss and the adaptation loss, and utilized information from both source dataset and target dataset to calibrate a shared model (i.e. the source network). After transferring layer-parameters from the source network to the target network, a series of fine-tuning strategies are conducted on the target network according to the target dataset.

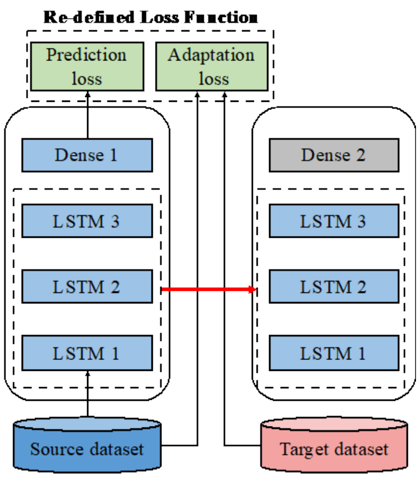

Figure 2. An example of the Self-adaptive Neural Network

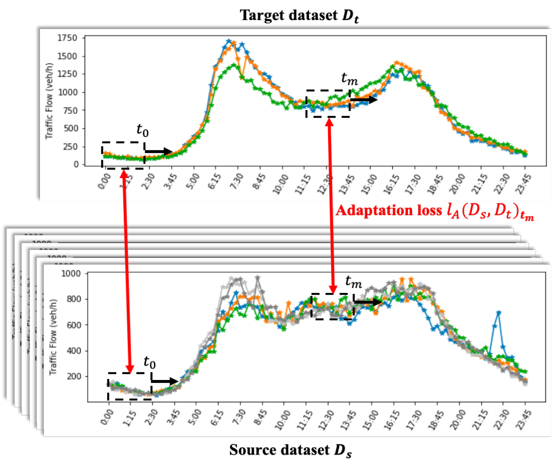

Figure 3. An example of the adaptation loss calculation

D. *Maximum Mean Discrepancy (MMD) measurement*

The Maximum Mean Discrepancy (MMD) is initially proposed in [20] to measure the distance between two sets of observations, which is very effective in finding samples that were generated from the same distribution. Given two probability distributions p and q, MMD is formulated as:

$$MMD(\mathcal{F}, p, q) := sup_{f\in\mathcal{F}}(\mathbf{E}_p[f(x)] - \mathbf{E}_q[f(y)]) \qquad (2)$$

where p and q are two Borel probability distributions, $\mathcal{F}$ is a class of functions $f: \mathcal{X} \to \mathbb{R}$.

Let $\left\{\mathbf{x}_s^{(i)}\right\}_{i=1,2,\dots,n_s}$ and $\left\{\mathbf{x}_t^{(j)}\right\}_{j=1,2,\dots,n_t}$ be samples that consist of independent and identically distributed observations from p and q, the empirical estimate of MMD can be defined as:

$$MMD(\mathbf{x}_s, \mathbf{x}_t) = \left\| \frac{1}{n_s}\sum_{i=1}^{n_s}\phi(\mathbf{x}_s^{(i)}) - \frac{1}{n_t}\sum_{j=1}^{n_t}\phi(\mathbf{x}_t^{(j)}) \right\|_{\mathcal{H}} \qquad (3)$$

where $\phi(\cdot)$ represents a feature space map to a universal reproducing kernel Hilbert space $\mathcal{H}$.

According to [19], the measurement can be rewritten as a kernelized equation form:

$$\begin{aligned} MMD^2(\mathbf{x}_s, \mathbf{x}_t) &= \frac{1}{n_s^2}\sum_{i=1}^{n_s}\sum_{j=1}^{n_s}k(\mathbf{x}_s^{(i)}, \mathbf{x}_s^{(j)}) + \frac{1}{n_t^2}\sum_{i=1}^{n_t}\sum_{j=1}^{n_t}k(\mathbf{x}_t^{(i)}, \mathbf{x}_t^{(j)}) \\ &- \frac{2}{n_s n_t}\sum_{i=1}^{n_s}\sum_{j=1}^{n_t}k(\mathbf{x}_s^{(i)}, \mathbf{x}_t^{(j)}) \\ &= \frac{\mathrm{Tr}(\mathbf{K}_{xss})}{n_s^2} + \frac{\mathrm{Tr}(\mathbf{K}_{xtt})}{n_t^2} - 2\frac{\mathrm{Tr}(\mathbf{K}_{xst})}{n_s n_t} \end{aligned} \qquad (4)$$

where $k(\mathbf{x}^{(i)}, \mathbf{x}^{(j)})$ represents all possible kernel functions.

The MMD measurement has been widely applied in transfer learning researches to evaluate the dataset differences and further reduce their distribution mismatch. Pan et al. [18] proposed the transfer component analysis (TCA) method to learn some transfer components across different domains in a reproducing kernel Hilbert space using MMD. Tzeng et al. [21] added an adaptation layer and a novel domain confusion loss (measured in MMD) in their CNN architecture to learn domain invariant features. More recently, some improved MMD measurements like Multi-Kernel MMD and Joint MMD are explored and applied in [16], [17] respectively to enhance the network adaptation capability.

In this paper, we attempt to incorporate the classic MMD measurement in the supervised learning process to achieve adaptive transfer learning in traffic prediction area. Other advanced loss function setting will be explored in our future study.

III. Case Study and Results

A. *Data description*

In our previous work [15], the experimental data are manually evaluated and selected according to their

Manuscript 269 submitted to 2020 IEEE Intelligent Transportation Systems Conference (ITSC). Received March 7, 2020.

geographical attributes, traffic demands and statistical distributions. To test the efficacy of our proposed Self-adaptive Neural Network in alleviating human efforts and computational burden in the data selection process, the source data are simply selected from adjacent links near the target link that suffers severe data missing problems. The cross-correlation coefficient is calculated to quantify the data patterns between datasets and provide different transfer scenarios to compare (e.g. transfer between consistent datasets or between inconsistent datasets).

The traffic data used in this paper are 15-minute interval Inductive Loop Detector (ILD) traffic flow obtained from the open source traffic data platform, Highways England Webtris (http://webtris.highwaysengland.co.uk/). Only four source links and one target link are selected to initially validate the efficacy of the proposed method. Specifically, the sensor sites M4/2188, M4/2302, M25/4963, M25/4854 are chosen as the source links and site M4/2240 is selected as the target link. For all the source links, the training data range from 1st January to 15th Sep 2019. For the target link, only 3-day data was chosen from 18th Sep 2019 to 20th Sep 2019 for training to artificially simulate the data insufficiency problem and the following 10-day data from 21st Sep 2019 to 30th Sep 2019 was chosen for testing. The cross-correlation coefficients between different sites are in Table 1 and the locations of selected sites are shown in Figure 4.

Table 1. Cross-correlation coefficients between datasets

Target link	Source links			
	M4/2188	M4/2302	M25/4963	M25/4854
M4/2240	0.9144	0.7988	0.8099	0.9069

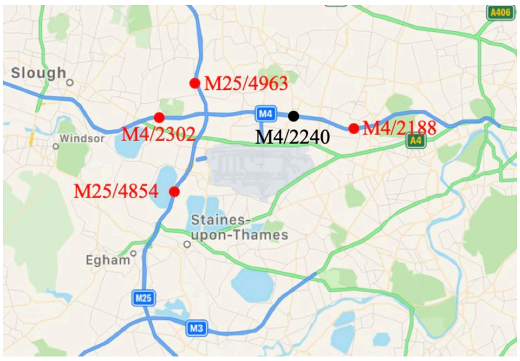

Figure 4. Locations of selected datasets

B. *Sliding window*

The sliding window is used in the data loading process to transform the original time sequence to multi-dimensional sequence, thus the traffic flow patterns can be adequately learned by the network. In this case, the input data and the label data are formulated as:

$$X_{input} = \begin{pmatrix} x_1 & \cdots & x_L \\ \vdots & \ddots & \vdots \\ x_K & \cdots & x_{K+L-1} \end{pmatrix} \quad X_{Label} = \begin{pmatrix} x_{L+1} \\ \vdots \\ x_{K+L} \end{pmatrix} \tag{5}$$

where $K = N + 1 - L$, L is the sequence length and N is the length of the original time sequence.

C. *Experiment setup*

Three groups of experiments are conducted in this section, namely baseline group, transfer learning group and adaptive transfer learning group. The baseline group only uses existing 3-day traffic flow data on the target link to calibrate the LSTM network and provide predictions for the following 10 days. The transfer learning group refers to the transfer framework in our previous study and the adaptive transfer learning group uses the proposed Self-adaptive Neural Network in this paper. According to our empirical studies and other related works [15], [17], [19], the hyperparameters in the case study are selected as Table 2. It is worth to note that our method redefines the loss function in the source network training process whereas most scholars in image classification area add the adaptation loss in the final fine-tuning process. Hence, the weight λ is positive in transfer learning in image classification but negative in our work. The fundamental idea is that the data inconsistency between the training data and the corresponding target data in each training batch can be detected and quantified in the adaptation loss function, if a batch of training data from the source dataset is highly inconsistent with the corresponding batch of the target dataset, a more negative adaptation loss will be added on the conventional prediction loss, resulting in a smaller overall loss. In this case, parameters in the network will be less influenced by the inconsistent source data thus the adaptive learning is achieved.

Table 2. Settings of the standard parameters

Leraning rate	0.02
Batch size	64
Iterations	1000
Hidden units	16
λ	-0.5
Transferred layers n	4
Sequence length L	5

For the MMD measurement, the Gaussian kernel is selected as the kernel function in this study since it has been accepted as a universal kernel and proven to make MMD useful in practice [22]. Therefore, the Kernel function in Equation (4) can be reformulated as the form in Equation (6). We expect the combination of MMD measurement and Gaussian kernel can provide accurate prediction results in adaptive transfer learning in short-term traffic prediction. The other distance measurements and Kernel functions will be investigated in our future studies.

$$k_{\text{Gaussian}}\left(\mathbf{x}^{(i)}, \mathbf{x}^{(j)}\right) = \exp\left(-\frac{\left\|\mathbf{x}^{(i)} - \mathbf{x}^{(j)}\right\|^2}{2\sigma^2}\right) \tag{6}$$

where σ is the standard deviation.

D. *Results and analysis*

The training process and initial experiment results of different models are shown in Table 3 and Figure 5,6. Only one-step ahead prediction (i.e. 15 minutes ahead) is conducted to verify the efficacy of our proposed network. The multi-step ahead experiments will be investigated in the future. Both unit dependent criteria (i.e. Mean Squared Error (MSE) and Mean Absolute Error (MAE)) and unit independent criteria (i.e.

Mean Absolute Percentage Error (MAPE)) are used to evaluate the prediction accuracy of different models. MAPE is more emphasized in this experiment since it provides a scale-independent and intuitive measure to compare.

Table 3. Experiment results

Model	Source link	Prediction error		
		MAE	MSE	MAPE
Baseline model	-	50.73	5298.3	10.11%
Transfer learning method	M4/2188	50.23	5196.55	8.86%
	M4/2302	52.09	5621.38	9.67%
	M25/4963	50.2	5201.6	10.06%
	M25/4854	48.09	4818.15	8.62%
Adaptive transfer learning method	M4/2188	48.19	4951.65	8.73%
	M4/2302	51.93	5471.24	9.24%
	M25/4963	50.22	5147.62	9.39%
	M25/4854	48.54	4843.2	8.77%

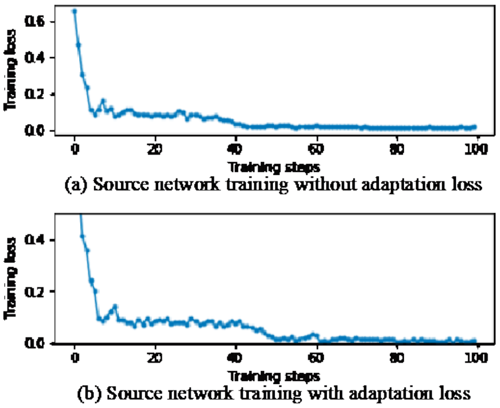

Figure 5. An example of training loss versus training steps

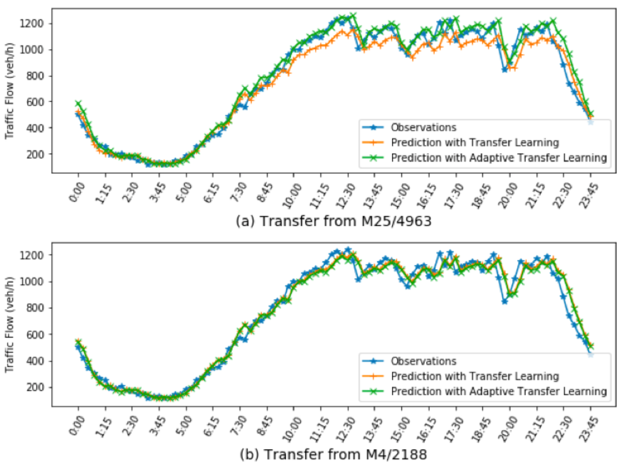

Figure 6. An example of different transfer scenarios

Table 3 summarizes experiment results of one-step ahead prediction with transfer learning and adaptive transfer learning under different transfer scenarios. The results show that both transfer learning and adaptive transfer learning methods are able to improve the prediction accuracy no matter the consistency between datasets. However, it is worth to note that the existing transfer learning method in [15] performs well with consistent source dataset (i.e. transfer from M4/2188 and M25/4854). However, the main limitation of the existing model is that it cannot predict properly when the source dataset is not consistent with the target dataset (i.e. transfer from M4/2302 and M25/4963). The results also show that it is essential to conduct the rigorous data selection process in the previous hybrid method, which significantly affects the final prediction results.

However, the proposed adaptive transfer learning method is less influenced by the consistency between datasets. Acceptable prediction accuracy can be achieved even with inconsistent source data from M4/2302 and M25/4963 (9.24% and 9.39% respectively), which is mainly due to the embedded adaptive learning process. As demonstrated in Figure 5, the training loss in our proposed Self-adaptive Neural Network fluctuate more and converge slower compared with the conventional training process without the adaptation loss (shown in Figure 5(b) and 5(a) respectively), which reveals the adaptive learning process.

An example of different transfer scenarios is shown in Figure 6. It can be seen in Figure 6, the two transfer learning methods perform similarly well when the highly consistent data from link M4/2188 is selected as the source dataset whereas the prediction accuracy of our previous method suffers more when the inconsistent source dataset from link M25/4963 is chosen. Based on this, there is evidence that this improved adaptive transfer learning method is able to adaptively filter inconsistent source data and learn target link related information in each training batch. Hence, the complex data selection process in our previous study can be alleviated.

IV. Conclusion and Future Work

This paper has proposed a novel short-term traffic prediction mdoel with an adaptive transfer learning method to alleviate the complex data selection process in previous transfer learning framework. Specifically, a Self-adaptive Neural Network with a novel domain adaptation loss is developed. The domain adaptation loss is able to calculate the distance between the source data and the corresponding target data in each training batch, thus the network can adaptively filter inconsistent source data and learn target link related information in each training batch. The Maximum Mean Discrepancy (MMD) measurement, which has been fully validated and applied in transfer learning research, is used in combination with Gaussian Kernel to measure the distance between the source data and target data in each training batch. A series of experiments are designed and conducted using 15-minute interval traffic flow data from Highways England. Based on the initial experiment results, there is evidence that the proposed adaptive transfer learning method is less influenced by the inconsistency between datasets and the rigorous data selection process can be alleviated. However, the difference in prediction accuracy between the two transfer learning methods is not significant and the conducted experiments are not sufficient for model validation. Below are some improvements which should be made in future studies:

- More experiments should be designed and tested to validate this method under different transfer scenarios. Especially, transfer between inconsistent datasets should be emphasized.

- Apart from adding an MMD measured domain adaptation loss to the overall loss function, some other potential adaptive learning techniques, like other Kernel functions, other distance evaluation criteria and the separate domain adaptation layer, should be studied.

- In this paper, all the layer parameters in the source network are transferred and finely tuned towards the target dataset. Although the proposed method is able to adaptively filter inconsistent source data and learn target link related information, some attributes extracted from the source dataset might be not helpful in the target dataset and negative transfer may appear in this case. As a result, this adaptive learning method should be used in combination with different transfer strategies to determine how many layer parameters should be transferred and to what extent the parameters should be fine-tuned.

ACKNOWLEDGMENT

The work is supported in part by the key program project by Ministry of Science and Technology, China, grant no. 2018YFB1600500. It is also supported by the Zhejiang University/University of Illinois at Urbana-Champaign Institute, and was led by Principal Supervisor Simon Hu.

REFERENCES

[1] Y. Kamarianakis and P. Prastacos, "Forecasting traffic flow conditions in an urban network: Comparison of multivariate and univariate approaches," Transp. Res. Rec., vol. 1857, (1), pp. 74-84, 2003.
[2] J. Yu et al, "Urban network-wide traffic speed estimation with massive ride-sourcing GPS traces," Transportation Research Part C: Emerging Technologies, vol. 112, pp. 136-152, 2020.
[3] X. Zhang and J. A. Rice, "Short-term travel time prediction," Transportation Research Part C: Emerging Technologies, vol. 11, (3-4), pp. 187-210, 2003.
[4] P. Duan et al, "STARIMA-based traffic prediction with time-varying lags," in 2016 IEEE 19th International Conference on Intelligent Transportation Systems (ITSC), 2016, .
[5] I. Okutani and Y. J. Stephanedes, "Dynamic prediction of traffic volume through Kalman filtering theory," Transportation Research Part B: Methodological, vol. 18, (1), pp. 1-11, 1984.
[6] F. Guo, J. W. Polak and R. Krishnamoorthy, "Predictor fusion for short-term traffic forecasting," Transportation Research Part C: Emerging Technologies, vol. 92, pp. 90-100, 2018.
[7] J. Guo et al, "Deep learning based congestion prediction using PROBE trajectory data," in Cictp 2019.
[8] B. Liao et al, "Deep sequence learning with auxiliary information for traffic prediction," in Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining, 2018.
[9] L. Zhou, S. Zhang, J. Yu and X. Chen, "Spatial-Temporal Deep Tensor Neural Networks for Large-Scale Urban Network Speed Prediction," in IEEE Transactions on Intelligent Transportation Systems.
[10] D. Kang, Y. Lv and Y. Chen, "Short-term traffic flow prediction with LSTM recurrent neural network," in 2017 IEEE 20th International Conference on Intelligent Transportation Systems (ITSC), 2017, .
[11] F. Guo, R. Krishnan and J. W. Polak, "Short-term traffic prediction under normal and incident conditions using singular spectrum analysis and the k-nearest neighbour method," IET and ITS Conference on Road Transport Information and Control (RTIC 2012), London, 2012, pp. 1-6.
[12] X. Ma et al, "Learning traffic as images: a deep convolutional neural network for large-scale transportation network speed prediction," Sensors, vol. 17, (4), pp. 818, 2017.
[13] J. Wang et al, "Traffic speed prediction and congestion source exploration: A deep learning method," in 2016 IEEE 16th International Conference on Data Mining (ICDM), 2016, .
[14] J. Guo et al, " GPS-based traffic congestion prediction using cnn-rnn and c3d hybrid model." in 98th Annual Meeting of Transportation Research Board, No. 19-05269, Washington D.C.
[15] J. Li et al, "Transfer learning in short-term traffic flow prediction with deep learning tools." in 99th Annual Meeting of Transportation Research Board, No. 20-01239, Washington D.C.
[16] M. Long et al, "Deep transfer learning with joint adaptation networks," in Proceedings of the 34th International Conference on Machine Learning-Volume 70, 2017, .
[17] M. Long et al, "Learning transferable features with deep adaptation networks," arXiv Preprint arXiv:1502.02791, 2015.
[18] S. J. Pan et al, "Domain adaptation via transfer component analysis," IEEE Trans. Neural Networks, vol. 22, (2), pp. 199-210, 2010.
[19] M. Ghifary, W. B. Kleijn and M. Zhang, "Domain adaptive neural networks for object recognition," in Pacific Rim International Conference on Artificial Intelligence, 2014, .
[20] K. M. Borgwardt et al, "Integrating structured biological data by kernel maximum mean discrepancy," Bioinformatics, vol. 22, (14), pp. e49-e57, 2006.
[21] E. Tzeng et al, "Deep domain confusion: Maximizing for domain invariance," arXiv Preprint arXiv:1412.3474, 2014.
[22] A. Gretton et al, "A kernel two-sample test," Journal of Machine Learning Research, vol. 13, (Mar), pp. 723-773, 2012.