Probabilistic Public Transport Demand Estimation with Graph Convolution Neural Network

Can Li[1], **Lei Bai**[1], **Wei Liu**[1,2], **Lina Yao**[1], **and S Travis Waller**[2]

[1]**School of Computer Science and Engineering, University of New South Wales, Sydney**
[2]**Research Centre for Integrated Transport Innovation, School of Civil and Environmental Engineering, University of New South Wales, Sydney**

Keywords: Graph Convolution Networks, Bayesian Inference, Probabilistic Prediction, Public Transport

INTRODUCTION

Public transport services serve a large number of people on a daily basis and play an irreplaceable role in metropolitan areas (Mouwen, 2015). The travel demand and mobility patterns are often the most critical inputs to aid efficient and effective planning and operation of the public transport systems. In particular, understanding, forecasting, and incorporating the demand uncertainty is necessary to optimally determine, e.g., bus fleet size, vehicle size, public transit lines and networks. Besides, uncertainties in public transport can result in bus bunching and unbalanced waiting flows across different stops, and appropriately forecasting and modeling demand uncertainty is fundamental to addressing these issues. However, most of the existing studies focused on deterministic demand estimation or mobility analytics while the confidence interval of the estimates has received much fewer attention (Hazelton, 2008).

This study proposes a Probabilistic Graph Convolution Model for capturing spatio-temporal correlations and providing travel demand estimation with a target confidence interval. Specifically, we formulate the Origin-Destination (OD) demand matrix on a graph and define each route between one OD pair as a node (in the context of neural network). A series of gated graph convolution layers are applied to capture the spatial and temporal correlations simultaneously instead of RNN-based (Recurrent Neural Network) methods which consume a larger number of iterative and accumulative errors in each future step Bai et al. (2019). We then utilize the Bayesian-based module to quantify the model uncertainty and obtain the passenger OD-demand interval with a specific level of confidence.

METHODOLOGY

The architecture of the proposed Probabilistic Graph Convolution Model (PGCM) is shown in Figure 1. It consists of two components: the Mobility Forecasting Module based on Graph-based Network and the Bayesian Approximation Module based on Monte Carlo dropout.

We first introduce the construction of the graph based on the OD-demand. OD-demand is not only related to location but also relies on demographic attributes and land use. Following Bai et al. (2019), we treat the city as a graph $\mathscr{G} = (V, E, A)$ and adopt gated mechanism for the graph convolution layers, where V is the set of routes, E is the set of edges and A is the adjacency matrix. In the graph, let $v_i \in V$ denotes a node and $e_{ij} = (v_i, v_j) \in E$ denotes an edge. We use Pearson Correlation Coefficient to measure the similarity $s_{ij} = Pearson(x_i(0-t), x_j(0-t))$ between the historical demand of two routes. The adjacency matrix $A \in \mathbf{R}^{N\times N}$ is determined by the demand similarity between two routes: $A_{ij} = \begin{cases} 1 & s_{ij} > \varepsilon \\ 0 & s_{ij} \leq \varepsilon \end{cases}$ where ε is the threshold to decide whether a correlation exists between two routs.

The result of graph convolution layer (gcl) can be calculated as: $X^{l+1} = (\tilde{D}^{-\frac{1}{2}}\tilde{A}\tilde{D}^{-\frac{1}{2}})X^l W$ where $\tilde{A} = A + I_n$, $\tilde{D}_{ii} = \sum_j \tilde{A}_{ij}$, W is the weight matrix of the layer. X^l denotes the input feature of demand in the l_{th} gcl. Then gating mechanism Dauphin et al. (2017) and residual learning He et al. (2016)

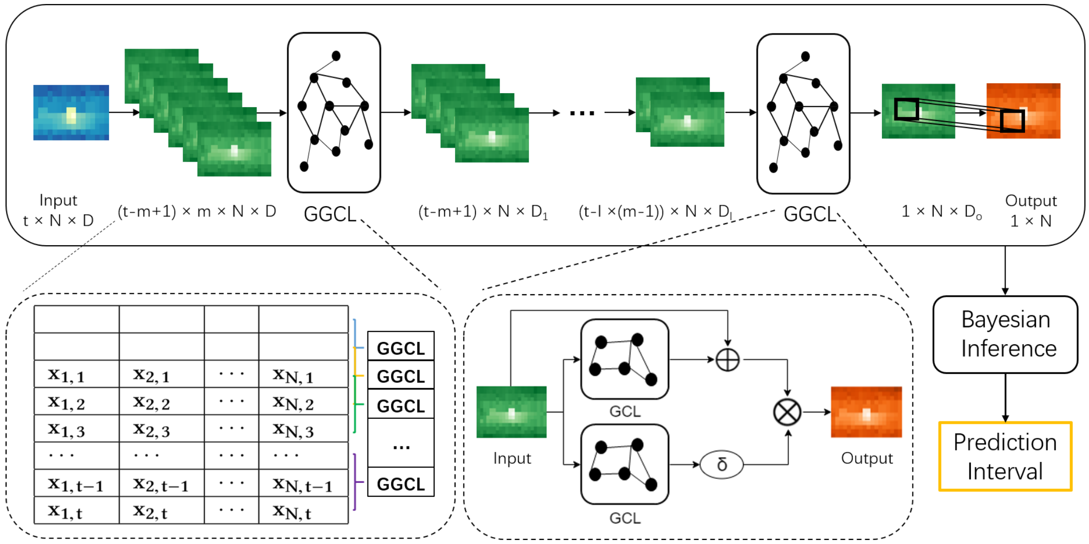

Figure 1. The Architecture of Probabilistic Graph Convolution Model (PGCM). $x_{i,j}$ represents the OD demand of j_{th} route in i_{th} time step. t denotes the used length of time steps, N denotes the number of routes, D denotes the dimension of the data, m denotes the used time steps in one gated graph convolution layer,δ represents the sigmoid function.

are adopted in the gcl to formulate the Graph Gated Convolution Layer (GGCL) which reduces the vanishing gradient problem. And the result of l_{th} layer is formulated as: $X^{l+1} = (\tilde{D}^{-\frac{1}{2}}\tilde{A}\tilde{D}^{-\frac{1}{2}})X^lW_1 + X^l) \otimes \delta(\tilde{D}^{-\frac{1}{2}}\tilde{A}\tilde{D}^{-\frac{1}{2}})X^lW_2)$ where δ denotes the sigmoid function and $\otimes$ represents the element-wise product operation.

Then we apply a series of GGCL along the temporal axis to capture the spatial-temporal correlations. We use the most recent t time steps historical OD demand to predict the demand in the next time step. In detail, we reshape the input demand $X \in \mathbf{R}^{t\times N\times D_{in}}$ into $(t-m+1)\times X_{new}$ where $X_{new} \in \mathbf{R}_{m\times N\times D_{in}}$ and m ($m < t$) is a hyperparameter to decide the time step length of one reshape input data. Then, each gated graph convolution layer analyzes N routes' spatial relations in m time steps to predict OD-demand. An illustration of data reshaping and spatial-temporal correlations capturing is shown in Figure 1. A set of GCCL series forms the module to capture the entire relations among t time steps. The output of l_{th} layer is $\hat{Y} \in \mathbf{R}^{t-l\times(m-1)\times N\times d_l}$. Therefore, only $\frac{t-1}{m-1}$ graph convolution layers are needed to capture the temporal dependencies which decrease the iterative steps and improves the accuracy compared with RNN-based models. The output of the set of GCCL is $\hat{Y} \in \mathbf{R}^{1\times N\times d_{out}}$ where d_{out} is the output dimension of the last GGCL. Finally, the output is sent into a convolution network for extracting dependencies among D_{out} dimensions' data.

In the training process of Mobility Forecasting Module, the objective is to minimize the error between the true OD-demand and the predicted values. The loss function is defined as the mean squared error for τ time steps formulated as: $L(\theta) = \sum_{i=T+1}^{T+\tau} ||\hat{X}_i - X_i||$ where θ denotes all the learnable parameters in the prediction model. It is solved via back-propagation and Adam optimizer.

Motivated by Gal and Ghahramani (2016), we adopt the Bayesian deep learning approach, which assumes that each weight and bias should obey a certain distribution instead of a certain value. This leads to an interval of the demand that results in a certain level of confidence. The construction of the network with dropout is not changed, so the prediction results are still reliable. Given a set of estimated values $\{\hat{X}_{T+1}, \hat{X}_{T+2}, \cdots, \hat{X}_{T+\tau},\}$ and true values $\{X_{T+1}, X_{T+2}, \cdots, X_{T+\tau}\}$, according to the Bayesian Theorem, the posterior distribution $P(W|X,\hat{X})$ is used to measure the probability of the parameters over the model. As suggested by Damianou and Lawrence (2013), the approximation of the Gaussian process is equivalent to a neural network with dropout. Thus, we use Gaussian distribution $q_\theta(W|X,\hat{X})$ to measure the posterior distribution of our model by minimizing the Kullback-Leibler divergence between them. Then the demand intervals are given by the Bayesian Inference. First, we randomly dropout some neural units with probability p before each layer, do the forward passes through the network, and get the prediction demand $\hat{X}^s$. Then, repeat the first step for S times, and we can get a set of predicted results $\{\hat{X}^1, \hat{X}^2, \cdots, \hat{X}^S\}$. At

Table 1. Comparison between the Proposed Method and Existing Methods

Method	RMSE	MAE	MAPE
ARIMA	29.6429	20.4692	0.6065
LR	28.5321	19.7717	0.5909
HA	13.4791	12.1329	0.4630
GCRN	17.6889	4.4272	0.5800
ConvLSTM	16.0453	6.4170	0.3303
GRU	10.1630	4.6386	0.2499
LSTnet	8.9854	**4.0135**	0.2515
Our	**7.8741**	**4.0412**	**0.2411**

Table 2. Comparison of Prediction Interval Performance

Level	Method	ConvLSTM	GRU	LSTNet	Our
94%	Span	9.3919	11.4451	8.6212	**7.0105**
	Proportion	28.18%	50.24%	46.43%	**56.41%**
95%	Span	9.7844	11.9545	8.9854	**7.3958**
	Proportion	29.26%	52.59%	49.09%	**57.91%**
96%	Span	10.2397	12.4937	9.4091	**7.6551**
	Proportion	30.58%	55.45%	52.59%	**59.66%**
97%	Span	10.8198	13.2014	9.9421	**8.0888**
	Proportion	32.11%	59.09%	57.03%	**61.72%**
98%	Span	11.5989	14.1520	10.6580	**8.6712**
	Proportion	34.12%	63.76%	62.73%	**64.27%**

last, the average value $\hat{\mu}$ and standard error $\hat{\eta}$ can be estimated to measure the prediction uncertainty: $\hat{\mu} = \bar{\hat{X}} = \frac{1}{S}\sum_{s=1}^{S}\hat{X}^s$ and $\hat{\eta}^2 = \frac{1}{S}\sum_{s=1}^{S}(\hat{X}^s - \hat{\mu})^2$. Therefore, with the help of Standard Normal Distribution table, we will get the prediction interval $[\hat{\mu} - z_{\alpha/2}\hat{\eta}, \hat{\mu} + z_{\alpha/2}\hat{\eta}]$ with a target level of confidence.

RESULTS

We compare the proposed method with some previous algorithms: Autoregressive Integrated Moving Average model (ARIMA), Linear Regression (LR), Historical Average (HA), Graph Convolutional Recurrent Network (GCRN) (Seo et al., 2018), Convolutional LSTM (ConvLSTM) (Xingjian et al., 2015), Gate Recurrent Unit (GRU) (Chung et al., 2014), and Long- and Short-term Time-series network (LSTNet) (Lai et al., 2018).

Table 1 summarizes the results of prediction accuracy for the proposed method and the listed tools. Although MAE of GCPM is slightly larger than that of LSTnet, RMSE and MAPE are far smaller than those for all other methods, which means that the proposed model has fewer big errors and provides more robust estimations. The results indicate that our model can capture spatio-temporal correlations more reliably estimate public transport mobility.

We also collect the results about the range of prediction interval with a certain level of confidence and the proportion of true values falling into the interval. Table 2 compares the results based on the proposed model and the other three strategies. The results show that the proposed model has a higher proportion while with a shorter prediction interval. This means that the proposed method produces more accurate and robust demand estimations.

CONCLUSION

This work provides confidence interval based OD-demand forecasting through exploring the relevance among temporal and spatial information of public transit data. We propose a Probabilistic Graph Convolution Model. In particular, we formulate the OD-demand on a graph and employ the Mobility Forecasting Module based on gated graph convolution layers to extract spatio-temporal correlations. The Bayesian Approximation Module is proposed to measure the model uncertainty and provide the confidence interval for the demand prediction. The results show that the proposed model outperforms other state-of-the-art methods. In the future, the current work will be extended by adding more information to the neural network to optimize the traffic flow or mobility prediction.

REFERENCES

Bai, L., Yao, L., Kanhere, S., Wang, X., Sheng, Q., et al. (2019). Stg2seq: Spatial-temporal graph to sequence model for multi-step passenger demand forecasting. *arXiv preprint arXiv:1905.10069*.

Chung, J., Gulcehre, C., Cho, K., and Bengio, Y. (2014). Empirical evaluation of gated recurrent neural networks on sequence modeling. *arXiv preprint arXiv:1412.3555*.

Damianou, A. and Lawrence, N. (2013). Deep gaussian processes. In *Artificial Intelligence and Statistics*, pages 207–215.

Dauphin, Y. N., Fan, A., Auli, M., and Grangier, D. (2017). Language modeling with gated convolutional networks. In *Proceedings of the 34th International Conference on Machine Learning-Volume 70*, pages 933–941. JMLR. org.

Gal, Y. and Ghahramani, Z. (2016). Dropout as a bayesian approximation: Representing model uncertainty in deep learning. In *international conference on machine learning*, pages 1050–1059.

Hazelton, M. L. (2008). Statistical inference for time varying origin–destination matrices. *Transportation Research Part B: Methodological*, 42(6):542–552.

He, K., Zhang, X., Ren, S., and Sun, J. (2016). Deep residual learning for image recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 770–778.

Lai, G., Chang, W.-C., Yang, Y., and Liu, H. (2018). Modeling long-and short-term temporal patterns with deep neural networks. In *The 41st International ACM SIGIR Conference on Research & Development in Information Retrieval*, pages 95–104. ACM.

Mouwen, A. (2015). Drivers of customer satisfaction with public transport services. *Transportation Research Part A: Policy and Practice*, 78:1–20.

Seo, Y., Defferrard, M., Vandergheynst, P., and Bresson, X. (2018). Structured sequence modeling with graph convolutional recurrent networks. In *International Conference on Neural Information Processing*, pages 362–373. Springer.

Xingjian, S., Chen, Z., Wang, H., Yeung, D.-Y., Wong, W.-K., and Woo, W.-c. (2015). Convolutional lstm network: A machine learning approach for precipitation nowcasting. In *Advances in neural information processing systems*, pages 802–810.