Abnormal metro passenger demand is predictable from alighting and boarding correlation

Zhanhong Cheng [1,2] Jiawei Wang [3] Martin Trépanier [4] Lijun Sun [1*]

[1]Department of Civil Engineering, McGill University [2]Department of Urban and Regional Planning, University of Florida
[3]Graduate School of Interdisciplinary Information Studies, University of Tokyo [4]Department of Mathematics and Industrial Engineering, Polytechnique Montreal

Introduction

Irregular sudden fluctuations in metro passenger demand during events or incidents can lead to critical supply or safety issues. However, forecasting abnormal metro passenger demand is challenging due to the absence of periodicity, high volatility, scarce samples, and the need for early warnings. We address abnormal metro passenger demand forecasting by leveraging the long-range Alighting-Boarding (AB) correlation driven by chained travel behavior. We found that leveraging the AB correlation enables early warnings of abnormal demand with up to a 5-hour lead time (depending on the activity duration), offering an effective abnormal demand warning solution that does not rely on auxiliary event data.

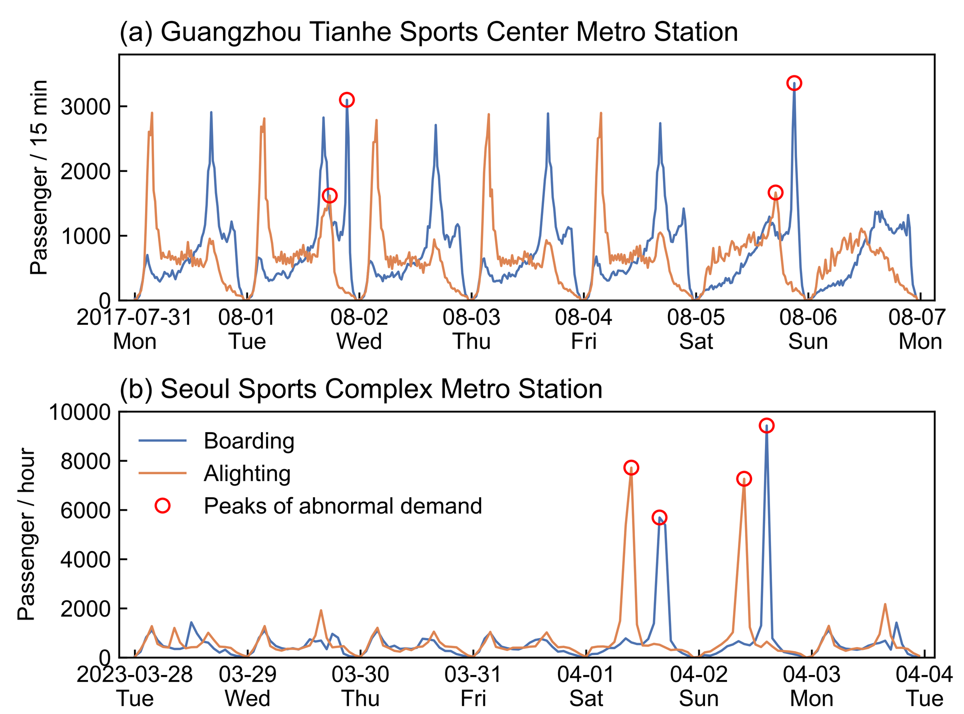

Figure 1. Passenger flow of two metro stations. Remarkably, a noticeable peak in abnormal alighting flow consistently precedes each peak in abnormal boarding flow.

Research highlights

- We argue that the alighting-boarding (AB) correlation should always be used for metro passenger demand forecasting. We show the salient effect of AB correlation in enhancing demand forecasting, particularly for early warming of abnormal boarding demand.
- We propose ABTransformer, a forecasting model that explicitly models the AB correlation with a bi-channel attention mechanism while maintaining explainability.
- We explore uncertainty quantification in metro demand forecasting, and demonstrate the multimodality of forecast distributions through clustering analysis.

Illustration of the alighting and boarding (AB) correlation

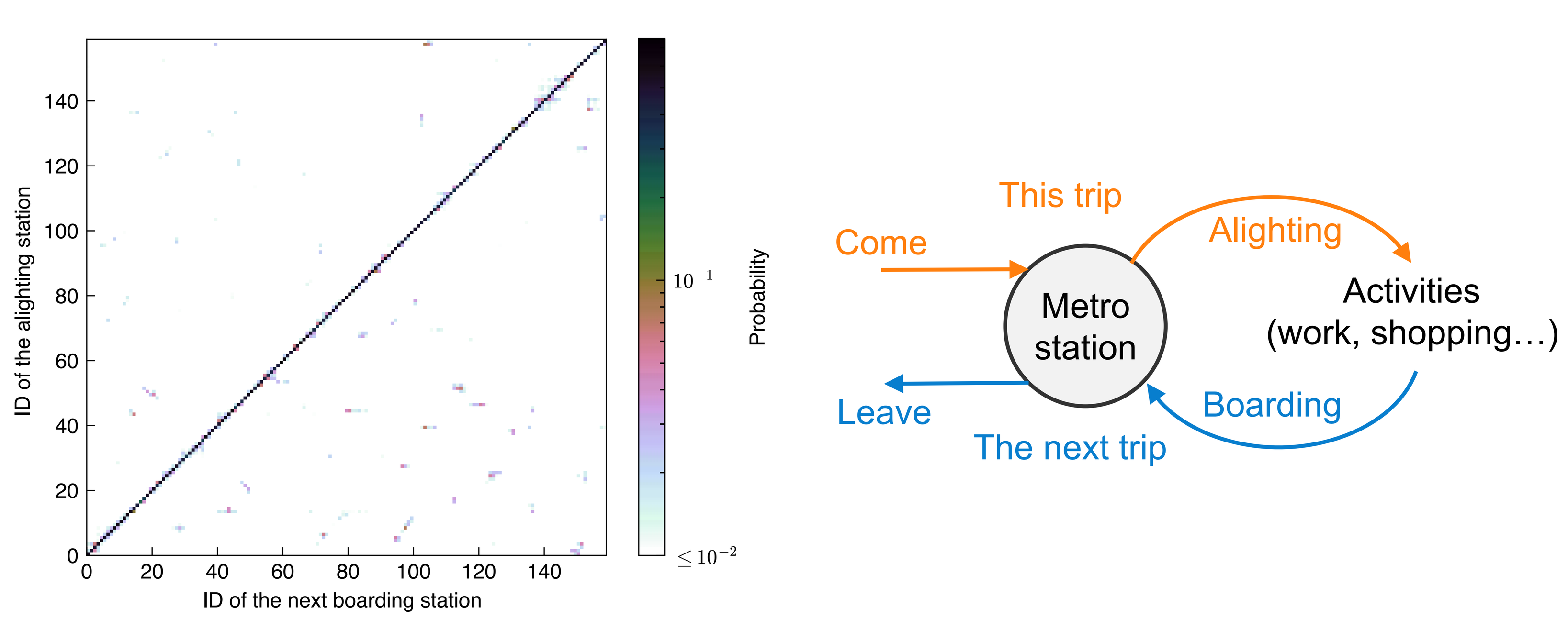

Figure 2. Illustration of alighting and boarding (AB) correlation. **Left**: The probability of the next boarding station given the current alighting station (Guangzhou metro data). **Right**: An illustration of chained behavior in metro systems.

The alighting-boarding Transformer (ABTransformer)

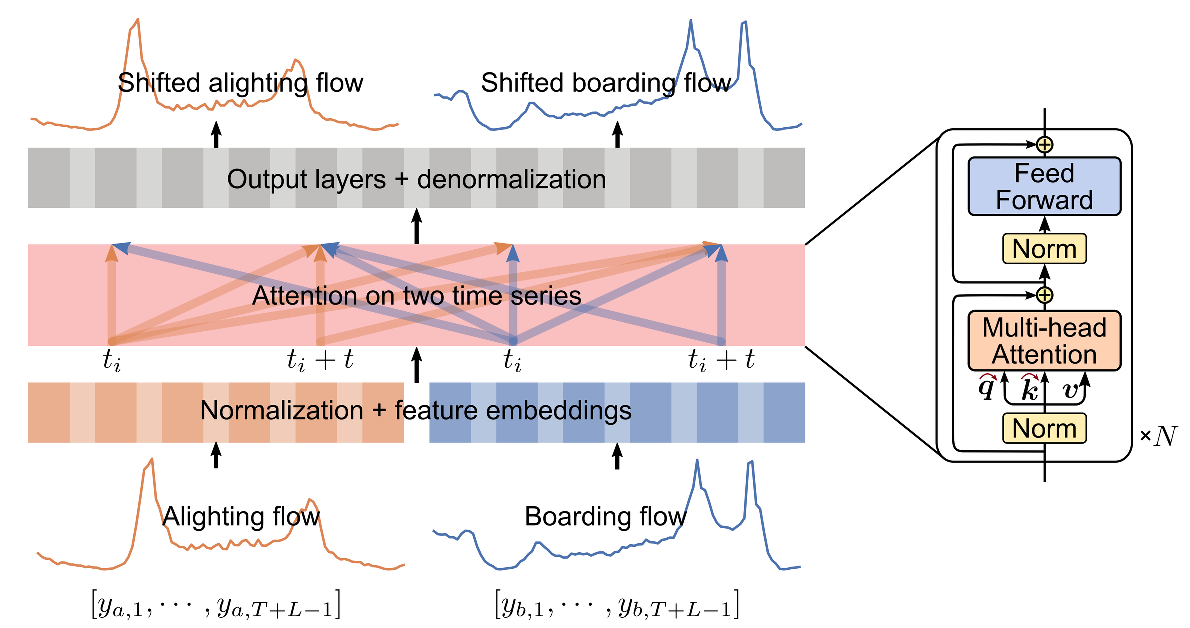

Figure 3. ABTransformer for metro passenger demand forecasting by explicitly modeling the correlations in the alighting and boarding flow.

- We use **bi-channel attention** to explicitly capture the AB correlation.

$$w_{i_1 i_2, t_1 t_2} = \frac{\exp\left(\frac{\boldsymbol{q}_{i_1,t_1}^\top \boldsymbol{k}_{i_2,t_2}}{\sqrt{d}} \boldsymbol{M}[t_1, t_2]\right)}{\sum_{i,t} \exp\left(\frac{\boldsymbol{q}_{i_1,t_1}^\top \boldsymbol{k}_{i,t}}{\sqrt{d}} \boldsymbol{M}[t_1, t]\right)}, \quad i_1, i_2, i \in \{a, b\}.$$

- The range of the AB correlation is determined by the activity duration (a relative time difference). Therefore, we use **rotary positional embeddings** (a type of relative positional embedding) in ABTransformer.

$$\boldsymbol{q}_{i,t} = \boldsymbol{R}^d_{\Theta,t} \boldsymbol{W}_q \boldsymbol{z}_{i,t}, \quad i \in \{a, b\}$$
$$\boldsymbol{k}_{i,t} = \boldsymbol{R}^d_{\Theta,t} \boldsymbol{W}_k \boldsymbol{z}_{i,t}, \quad i \in \{a, b\}$$

where $\boldsymbol{R}^d_{\Theta,t}$ is the the rotary matrix. Su et al. 2024 showed that the dot product of a query and a key at any two time points t_1, t_2 after the rotation is a function that contains the relative position information between the two time steps :

$$\boldsymbol{q}_{i_1,t_1}^\top \boldsymbol{k}_{i_2,t_2} = f\left(\boldsymbol{z}_{i_1,t_1}, \boldsymbol{z}_{i_2,t_2}, t_1 - t_2\right), \quad i_1, i_2 \in \{a, b\}.$$

Interpreting the forecasting with attention weights

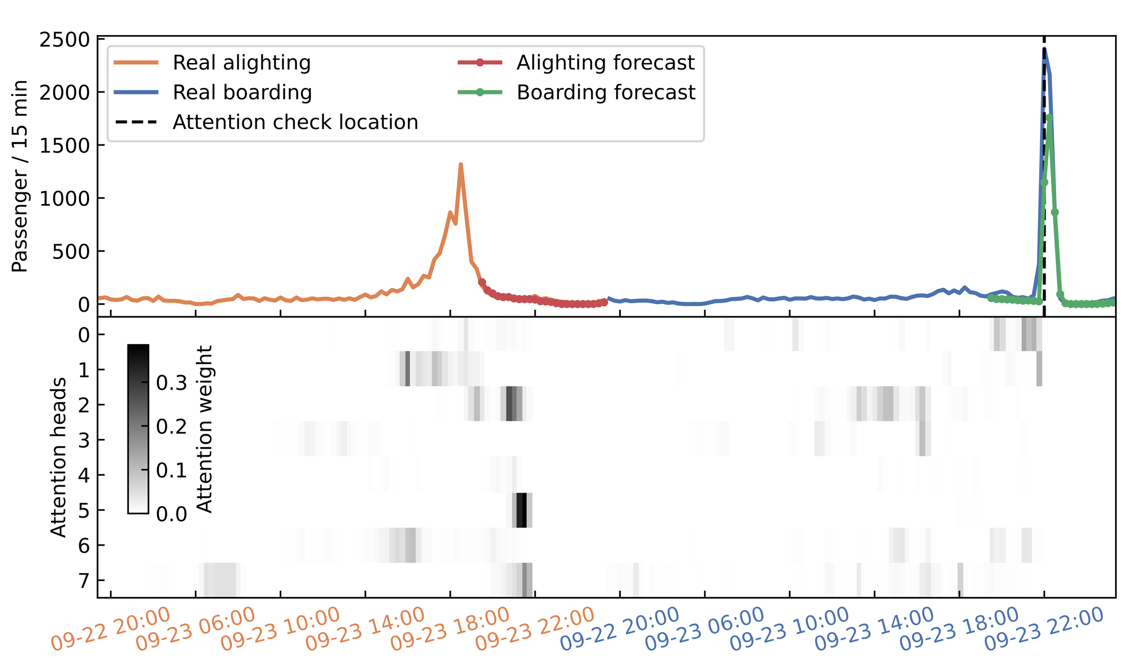

Figure 4. Interpreting boarding demand forecasting in a metro station in Guangzhou using attention weights of the last layer in the ABTransformer. The boarding demand forecast at the checked location exhibits significant attention to periods with abnormal alighting demand, indicating the parts of the input sequence that contribute to the forecast at the checked location.

Forecasting performance

Table 1. Deterministic forecast performance. N: normal flow, A: abnormal flow, T: total flow. Subscripts on model names indicate the forecasting approach used: (I) independent, (J1) joint approach 1, or (J2) joint approach 2. The best performance is highlighted in bold.

Metric	Model	Guangzhou						Seoul					
		Boarding			Alighting			Boarding			Alighting		
		N	A	T	N	A	T	N	A	T	N	A	T
MAE	DeepAR$_I$	48.8	428.1	49.0	43.6	354.3	43.9	48.4	483.6	50.2	51.8	473.2	53.0
	DeepAR$_{J1}$	43.1	393.6	43.3	43.6	343.3	43.7	**44.6**	435.0	**46.1**	**51.2**	467.3	**52.4**
	Nlinear$_I$	67.3	417.7	67.6	68.0	357.9	68.1	150.4	564.2	152.0	152.9	525.3	153.9
	Nlinear$_J$	63.8	393.4	64.0	64.6	363.6	64.6	125.5	453.3	126.8	134.7	513.1	135.8
	Transformer$_I$	48.8	424.8	49.1	46.5	363.0	46.5	52.1	488.8	53.8	54.8	461.1	56.0
	Transformer$_{J1}$	**40.2**	366.6	**40.4**	**40.5**	**319.5**	**40.6**	47.2	428.9	48.7	52.1	**442.1**	53.2
	Transformer$_{J2}$	41.7	**352.1**	41.9	42.8	348.1	42.9	48.5	**404.5**	49.9	53.1	458.1	54.2
	Transformer$^*_{J2}$	41.9	379.8	42.2	43.1	340.1	43.3	46.7	427.2	48.2	55.1	492.7	56.3
WMAPE	DeepAR$_I$	12.1%	45.0%	12.2%	10.4%	37.0%	10.4%	6.0%	36.4%	6.2%	6.4%	36.6%	6.6%
	DeepAR$_{J1}$	10.7%	41.4%	10.8%	10.9%	35.5%	10.9%	**5.5%**	33.1%	**5.7%**	**6.3%**	36.4%	**6.5%**
	Nlinear$_I$	16.7%	43.9%	16.8%	16.9%	37.0%	16.9%	18.7%	42.5%	18.8%	19.0%	40.6%	19.1%
	Nlinear$_J$	15.9%	41.4%	15.9%	16.1%	37.6%	16.1%	15.6%	34.1%	15.7%	16.7%	39.6%	16.8%
	Transformer$_I$	12.2%	44.7%	12.2%	11.6%	37.6%	11.6%	6.5%	36.8%	6.7%	6.8%	35.6%	6.9%
	Transformer$_{J1}$	**10.0%**	38.6%	**10.0%**	**10.1%**	**33.0%**	**10.1%**	5.9%	32.4%	6.0%	6.5%	**34.4%**	6.6%
	Transformer$_{J2}$	10.4%	**37.0%**	10.4%	10.7%	36.0%	10.7%	6.0%	**30.5%**	6.2%	6.6%	35.4%	6.7%
	Transformer$^*_{J2}$	10.4%	39.9%	10.5%	10.7%	36.3%	10.8%	6.2%	35.4%	6.2%	6.9%	37.5%	7.0%

* Using absolute positional embeddings instead of rotary positional embeddings.

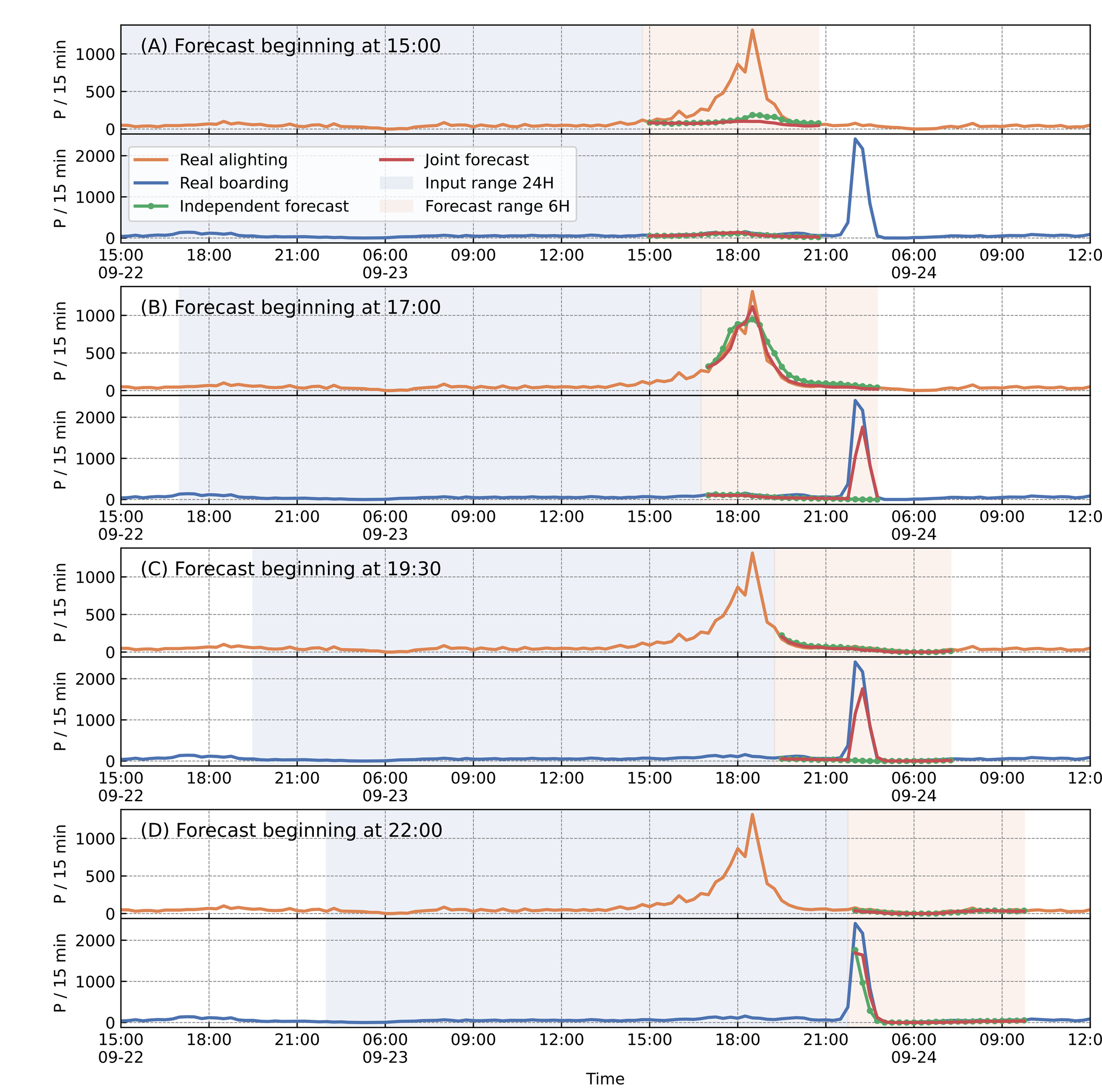

Figure 5. Comparison of early warning capabilities for abnormal boarding demand: independent model (Transformer$_I$) vs. joint model (Transformer$_{J2}$).

Acknowledgement

This research is supported by the Natural Sciences and Engineering Research Council (NSERC) of Canada. Z. Cheng would like to thank Dr. Shenhao Wang from the University of Florida for supporting the conference travel.

Scan for Preprint