東京大学　関本研究室 / Sekimoto Lab. IIS, The University of Tokyo

# Spatio Heterogeneity Mixture-of-Experts Embedding for Traffic Forecasting

Hangchen Liu, Renhe Jiang, Yoshihide Sekimoto

## Background

With the rapid development of the Intelligent Transportation System, accurate traffic forecasting has emerged as a critical challenge. Given the traffic series $X_{t|T+1}$:t in the previous T time frames, traffic forecasting aims to infer the traffic data in the future T^′ frames by training a model F(·) with parameters θ , which can be formulated as :

$$[X_{t-T+1}, \ldots, X_t] \xrightarrow{F(\cdot),\ \theta} [X_{t+1}, \ldots, X_{t+}$$

The advancements in recent network architectures have encountered diminishing performance gains, prompting a shift in focus from complicated model designs towards effective representation techniques for the data itself. In light of this, we focus on input embedding, a widely-used, simple, yet powerful representation technique, that is often overlooked in terms of its effectiveness.

## Model

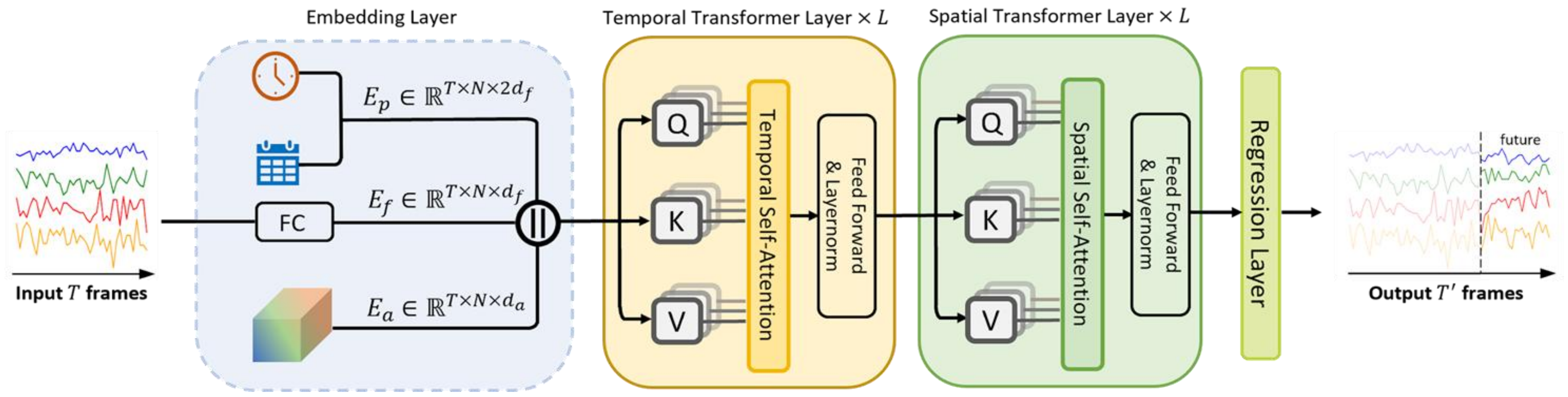


Figure 1. Model Architecture for SHMoE

## Performance

| Dataset | PEMS04 | | | PEMS07 | | | PEMS08 | | |
|---|---|---|---|---|---|---|---|---|---|
| Metric | MAE | RMSE | MAPE | MAE | RMSE | MAPE | MAE | RMSE | MAPE |
| GWNet | 18.53 | **29.92** | 12.89% | 20.47 | 33.47 | 8.61% | 14.40 | 23.39 | 9.21% |
| DCRNN | 19.63 | 31.26 | 13.59% | 21.16 | 34.14 | 9.02% | 15.22 | 24.17 | 10.21% |
| AGCRN | 19.38 | 31.25 | 13.40% | 20.57 | 34.40 | 8.74% | 15.32 | 24.41 | 10.03% |
| STGCN | 19.57 | 31.38 | 13.44% | 21.74 | 35.27 | 9.24% | 16.08 | 25.39 | 10.60% |
| GTS | 20.96 | 32.95 | 14.66% | 22.15 | 35.10 | 9.38% | 16.49 | 26.08 | 10.54% |
| MTGNN | 19.17 | 31.70 | 13.37% | 20.89 | 34.06 | 9.00% | 15.18 | 24.24 | 10.20% |
| STNorm | 18.96 | 30.98 | 12.69% | 20.50 | 34.66 | 8.75% | 15.41 | 24.77 | 9.76% |
| GMAN | 19.14 | 31.60 | 13.19% | 20.97 | 34.10 | 9.05% | 15.31 | 24.92 | 10.13% |
| PDFormer | 18.36 | 30.03 | 12.00% | 19.97 | 32.95 | 8.55% | 13.58 | 23.41 | 9.05% |
| STID | 18.38 | 29.95 | 12.04% | 19.61 | 32.79 | 8.30% | 14.21 | 23.28 | 9.27% |
| **STAEformer** | 18.22 | 30.18 | **11.98%** | 19.14 | 32.60 | **8.01%** | 13.46 | 23.25 | 8.88% |
| Ours | **18.17** | 30.01 | 11.99% | **19.05** | **32.42** | 8.03% | **13.38** | **22.97** | **8.84%** |

Table 1. Performance of SHMoE

## Dataset

Our method is verified on five traffic forecasting benchmarks, i.e., METR-LA, PEMS-BAY, PEMS04, PEMS07, and PEMS08. The first two datasets were proposed by previous works in this field. The time interval in the six datasets is 5 minutes, so there are 12 frames in each hour. More details are shown in Table 2

| Dataset | #Sensors (N) | #Timesteps | Time Range |
|---|---|---|---|
| METR-LA | 207 | 34,272 | 03/2012 - 06/2012 |
| PEMS-BAY | 325 | 52,116 | 01/2017 - 05/2017 |
| PEMS04 | 307 | 16,992 | 01/2018 - 02/2018 |
| PEMS07 | 883 | 28,224 | 05/2017 - 08/2017 |
| PEMS08 | 170 | 17,856 | 07/2016 - 08/2016 |

Table 2. Summary of Datasets

## Case Study

SHMoE achieves better performance on most metrics on all five datasets without any graph modeling. The encouraging results indicate that STHMoE shows our model can capture different patterns while similar input sequence on different spatial nodes. Since we decoupled the input similarity and the output feature similarity, it brings our model the ability to handle with spatial heterogeneity.

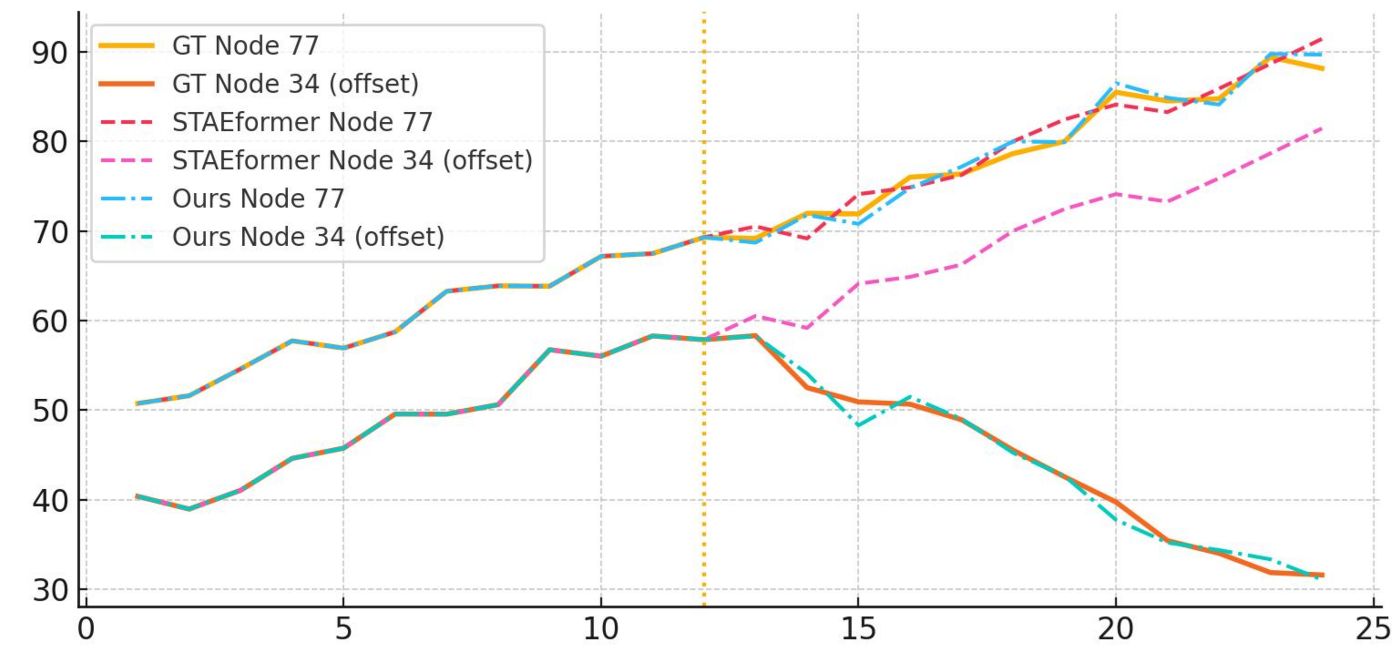


Figure 2. Predictions on METRLA for STAEformer and Ours

## Conclusion

In this study, we focus on a basic representation learning technique for traffic time series forecasting, i.e., input embedding. We propose a novel spatio heterogeneity mixture-of-experts that can work on vanilla transformers to achieve the SOTA performance on five traffic benchmarks. Further studies demonstrate that our model can effectively capture intrinsic spatio-temporal dependencies. Instead of designing complicated models, our study shows a promising direction for addressing the challenges in traffic forecasting.