

박사학위논문
Ph.D. Dissertation

도시 인간 활동 이해를 반영한 시공간 교통 예보
모델의 개선

Enhancing Spatiotemporal Traffic Forecasting Models with
Insights from Urban Human Activity

2024

한수민 (韓秀岷 Han, Sumin)

한국과학기술원

Korea Advanced Institute of Science and Technology

박사학위논문

도시 인간 활동 이해를 반영한 시공간 교통 예보
모델의 개선

2024

한수민

한국과학기술원

전산학부

도시 인간 활동 이해를 반영한 시공간 교통 예보 모델의 개선

한 수 민

위 논문은 한국과학기술원 박사학위논문으로
학위논문 심사위원회의 심사를 통과하였음

2024년 11월 26일

심사위원장 이 동 만 
심사위원 차 미 영 
심사위원 김 승 겸 
심사위원 김 태 균 
심사위원 안 지 선 

Enhancing Spatiotemporal Traffic Forecasting Models with Insights from Urban Human Activity

Sumin Han

Advisor: Dongman Lee

A dissertation submitted to the faculty of
Korea Advanced Institute of Science and Technology in
partial fulfillment of the requirements for the degree of
Doctor of Philosophy in Computer Science

Daejeon, Korea
December 18, 2024

Approved by



Dongman Lee
Professor of Computer Science

The study was conducted in accordance with Code of Research Ethics¹.

¹ Declaration of Ethical Conduct in Research: I, as a graduate student of Korea Advanced Institute of Science and Technology, hereby declare that I have not committed any act that may damage the credibility of my research. This includes, but is not limited to, falsification, thesis written by someone else, distortion of research findings, and plagiarism. I confirm that my thesis contains honest conclusions based on my own careful research under the guidance of my advisor.

DCS

한수민. 도시 인간 활동 이해를 반영한 시공간 교통 예보 모델의 개선.
전산학부 . 2024년. 53+iv 쪽. 지도교수: 이동만. (영문 논문)

Sumin Han. Enhancing Spatiotemporal Traffic Forecasting Models with Insights from Urban Human Activity. School of Computing . 2024.
53+iv pages. Advisor: Dongman Lee. (Text in English)

초 록

교통 예보는 시민의 안전과 편의를 보장하는 중요한 요소 중 하나이다. 기존의 교통 예보 모델은 주로 공간적 및 시간적 상관관계를 포착하기 위해 딥러닝 구조에 중점을 두고 있다. 그러나 이러한 모델은 교통의 본질적인 특성을 간과하는 경우가 많다. 특히 대부분의 교통 데이터셋에서 사용되는 센서 네트워크는 실제 차량이 사용하는 도로 네트워크를 정확히 반영하지 않아 도시 활동에서 발생하는 교통 패턴에 대한 통찰을 제공하지 못한다. 이러한 한계를 극복하기 위해 본 논문에서는 그래프 컨볼루션 딥러닝 알고리즘을 기반으로 한 개선된 교통 예보 방법을 제안한다. 가구주택통행조사에서 수집한 인간 활동 빈도 데이터를 활용하여 활동과 교통 패턴 간 인과 관계를 추론하는 능력을 향상시켰다. 기존의 그래프 컨볼루션 순환 네트워크 및 그래프 컨볼루션 트랜스포머 구조에 최소한의 수정만 가했음에도 불구하고, 본 접근법은 과도한 계산 부담 없이 최첨단 성능을 달성하였다.

핵심 낱말 교통 예보, 그래프 컨볼루션 딥러닝, 인간 활동 데이터

Abstract

Traffic forecasting is one of the key elements to ensure the safety and convenience of citizens. Existing traffic forecasting models primarily focus on deep learning architectures to capture spatial and temporal correlation. They often overlook the underlying nature of traffic. Specifically, the sensor networks in most traffic datasets do not accurately represent the actual road network exploited by vehicles, failing to provide insights into the traffic patterns in urban activities. To overcome these limitations, we propose an improved traffic forecasting method based on graph convolution deep learning algorithms. We leverage human activity frequency data from National Household Travel Survey to enhance the inference capability of a causal relationship between activity and traffic patterns. Despite making minimal modifications to the conventional graph convolutional recurrent networks and graph convolutional transformer architectures, our approach achieves state-of-the-art performance without introducing excessive computational overhead.

Keywords Traffic Forecasting, Graph Convolution Deep Learning, Human Activity Data

Contents

Contents	i
List of Tables	iii
List of Figures	iv
Chapter 1. Introduction	6
1.1 Overview	6
1.2 Challenge	8
1.2.1 Construction of an Accurate Sensor Adjacency Network	8
1.2.2 Addressing Individual Sensor Spatial Heterogeneity . .	9
1.2.3 Integration of Urban Human Activity	10
1.3 Contribution	11
Chapter 2. Literature Review	12
2.1 Spatiotemporal Deep Learning Traffic Forecasting Models . . .	12
2.1.1 Proposal of the Traffic Forecasting Problem	12
2.1.2 Traffic Datasets	12
2.1.3 Diffusion Convolutional Recurrent Neural Network . . .	13
2.2 Advanced Spatiotemporal Traffic Forecasting Models	15
2.3 Human-Centered Understanding of Urban Mobility	20
2.3.1 Traffic Wave Theory	20
2.3.2 Traffic Speed-Volume Relationship	21
2.3.3 Activity-Based Household Travel Surveys	22
2.3.4 Origin-Destination Based Movement Behavior	24
Chapter 3. Human-Activity-based Traffic Forecasting Models	26
3.1 Problem Formulation	26
3.2 Methodology	26
3.2.1 Graph Construction	26
3.2.2 Sensor Embedding	28
3.2.3 Activity Embedding	28
3.2.4 Deep Neural Network	29
3.3 Experimental Setting	30
3.3.1 Data Description and Preprocessing	30
3.3.2 Evaluation Setup	32
3.4 Results	36

3.4.1	Performance Comparison	36
3.4.2	Ablation Study	40
3.4.3	Case Study	42
3.4.4	Sensor Reactions Based on Activity Input	44
Chapter 4.	Discussion	45
4.1	Discussion on Traffic Forecasting	45
4.1.1	About UAGCRN	45
4.1.2	Challenges in Applying UAGCRN to South Korea . . .	45
4.2	Thoughts on True Future Prediction	47
Chapter 5.	Conclusion	48
Bibliography		49
Acknowledgments in Korean		52
Curriculum Vitae		53

List of Tables

3.1	Data statistics (B.C.: Normalized Betweenness Centrality). *PEMSD7 only contains weekdays.	31
3.2	Forecasting error in METR-LA, PEMS-BAY, PEMSD7 datasets. \dagger represents the model leveraging our co-occurrence and distance-based adjacency matrix. * represents the model self-trains the sensor adjacency. Best and second best results are represented as BOLD and <u>underline</u> .	35
3.3	Computational cost of METR-LA under the same environment. The number of stacks is $L = 5$ in GMAN and $L = 3$ in UAGCTF \dagger , while DCRNN, UAGCRN \dagger do not have stacked architecture ($L = 1$).	40
3.4	Ablation study of UAGCRN \dagger and UAGCTF \dagger by replacing AE with timestamp embedding (TE). Best and second best results are represented as BOLD and <u>underline</u> .	42

List of Figures

1.1	A Glimpse into the Short-term Traffic Forecasting Challenge.	6
1.2	(a) Traffic Congestion as Depicted in the SimCity 2013 (b) Forrester’s Global System Dynamics Diagram (1971)	7
1.3	Problem of Data-driven Trainable Sensor Adjacency Models.	9
1.4	Histogram Representing Sensor Heterogeneity in Speed Values.	10
1.5	Daily Human Activity of a Person	10
1.6	Temporal Encodings leveraged in GMAN	11
2.1	(a) METR-LA and PEMS-BAY from DCRNN (b) PeMSD7 (M, L) from STGCN	13
2.2	Sequence to Sequence Encoder-Decoder of DCRNN	14
2.3	Mainstream Traffic Forecasting Models, with our model (UAGCRN) highlighted in purple.	17
2.4	Implemented Spatiotemporal models in LibCity	19
2.5	Bidirectional Traffic Waves	20
2.6	(a) Nonlinear Relationship Between Traffic Speed and Traffic Volume (b) Empty Road vs. Congested Road	21
2.7	National Household Travel Survey Webpage	22
2.8	Sample of Seoul Floating Population Movement Data	23
2.9	Graph of Floating Population Movement Data in Seoul by Day and Hour	23
2.10	(a) Random Walk (b) Levy Flight (c) Trapline Foraging	24
2.11	Node2Vec Illustration	24
3.1	The A* algorithm is utilized to generate travel paths between the origin (red) and the destination (orange), sampled from a grid pair, with different costs of using the freeway: the ideal shortest path (blue), and paths that make greater use of the freeway (pink, cyan). The gray markers represent the traffic sensors, which do not necessarily appear on the generated travel paths (in METR-LA).	27
3.2	(a) Stacked visualization of generated travel paths. (darker color – frequency of the sensor appearance) (b) A travel path contains OSM node IDs and sensor IDs like a sentence.	28
3.3	Urban human activity frequencies from the National Household Travel Survey for Activity Embedding.	29
3.4	Model Architecture (UA-GCRN)	29
3.5	Model Architecture (UA-GCTransformer)	33
3.6	Traffic sensors (red markers) along with OSM freeways (blue paths) and the corresponding freeways where the traffic sensors are located (green paths) in PEMS-BAY. The partitioned grid is also represented with dark green squares.	34
3.7	Adjacency Matrix Visualization: Legacy, Co-Occurrence Adjacency, and Final Graph in PE MSD7.	34
3.8	METR-LA (UA-GCRN, UA-GCTransformer)	37
3.9	METR-LA (UA-LSTM, UA-Transformer)	37
3.10	Ablation Test (RMSE) of our modules – Our Graph(G), SE, A on METR-LA.	37

3.11 PEMS-BAY (UA-GCRN, UA-GCTransformer)	38
3.12 PEMS-BAY (UA-LSTM, UA-Transformer)	38
3.13 Ablation Test (RMSE) of our modules – Our Graph(G), SE, A on PEMS-BAY.	38
3.14 PEMSD7 (UA-GCRN, UA-GCTransformer)	39
3.15 PEMSD7 (UA-LSTM, UA-Transformer)	39
3.16 Ablation Test (RMSE) of our modules – Our Graph(G), SE, A on PEMSD7.	39
3.17 Performance degradation in UADGRU \dagger as the number of diffusion steps (K) increases. .	41
3.18 METR-LA, outperforms with our graph	43
3.19 PEMS-BAY, outperforms with our graph	43
3.20 Sensor Reactions Based on Activity Information with UAGCRN (Red/Green: more/less congestion)	44

THE CITIES OF THE FUTURE WILL NO LONGER BE COMPLETED SOLELY BY HUMAN EFFORTS.

T Artificial intelligence and robots will replace humans in repetitive labor. When designing new structures, AI can propose creative architectural designs according to given conditions, and during construction, robots can take over repetitive and physical tasks such as moving bricks or applying cement. Let us call these AI and robotic systems developed to support urban architecture and construction as “Urban Space Robots.”

As Urban Space Robots are efficiently utilized in the future, the role and responsibility of urban planners, who are the decision-makers using these systems, will become more significant. In the past, it took a considerable amount of time from design to construction, but thanks to advancements in Urban Space Robots, a time has come when urban planning in simulations is quickly realized in reality. However, if decision-makers abuse Urban Space Robots for personal power and gain, the consequences would be devastating. Human rights will be ignored, the environment destroyed, dilapidated abandoned buildings would proliferate, crime and fire incidents would surge, and poor financial management might make it difficult to provide even basic welfare. Additionally, just as social media penetrated deeply into people’s minds, controlling and deteriorating mental well-being, Urban Space Robots will pose a dangerous risk by penetrating the physical world of humans, even controlling bodily movement.

Currently, due to urbanization, we are facing various problems. With urban populations highly concentrated, numerous issues such as skyrocketing real estate prices, declining value of labor income, the spread of speculative get-rich-quick attitudes, declining marriage rates and low birth rates, rising individualism, and increasing solitary deaths have arisen. These issues may be less a matter of individual responsibility and more a form of “cost” our society is paying for advancing cities in pursuit of short-term economic profit without sufficient discussion. Furthermore, historically, even individuals who loved their neighbors and were dedicated to the nation for a better city have been attacked or disparaged due to jealousy and conflicts of interest, and sometimes their contributions were undervalued or distorted historically. As a result, even those with both power and goodwill have become hesitant to take on leadership roles.

In the future, Urban Space Robots will connect more deeply with citizens’ communications. Citizens will actively express the discomfort, anxiety, and dissatisfaction they feel in their daily lives, and AI will aggregate this information through big data collection devices, reflecting it in urban development. This process will provide opportunities to reduce the unhappiness of citizens, understand each other’s positions and roles, and communicate more effectively. We look forward to a future where, instead of a small group of leaders unilaterally leading the way, every citizen contributes to a better city and community from their own place, guiding the development of the nation and humanity together. Through Urban Space Robots, we hope to enhance communication and neighborly love, and that citizens will understand that leaders, too, are humans like us, working together toward a happier city. This dissertation aims to make a small footprint toward that future.

October 25th, 2024

Daejeon

For whoever wants to save their life will lose it,
but whoever loses their life for me will save it.

Luke 9:24

Chapter 1. Introduction

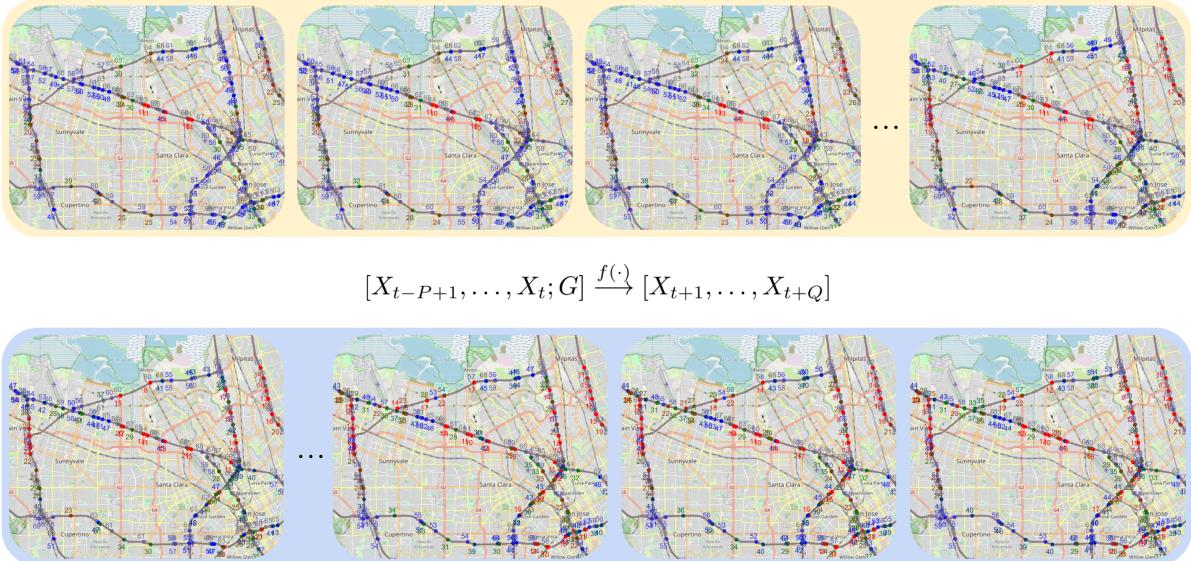


Figure 1.1: A Glimpse into the Short-term Traffic Forecasting Challenge.

1.1 Overview

The **Traffic Forecasting**¹ problem addressed in this dissertation is defined as predicting future traffic values by measuring traffic speed or volume from highway-installed traffic sensors or cameras. Figure 1.1 provides an intuitive depiction of the Traffic Forecasting problem, where past traffic scenes within the yellow range are used to forecast future traffic scenes in the blue range². Efficient Traffic Forecasting can be integrated into navigation systems to help select routes that avoid traffic congestion, thereby saving time for personal vehicles or logistics transport vehicles. Furthermore, integrating Traffic Forecasting technology with future autonomous vehicle and platooning technologies³ could enable efficient traffic management by strategically allocating vehicle routes and numbers from a holistic perspective.

Fundamentally, the most intriguing aspect of the Traffic Forecasting problem is imagining how traffic would change in response to unforeseen urban transformations, such as the construction of new residential buildings or large-scale employment facilities. For instance, if high-density residential areas were developed in a specific region to accommodate a newly created workforce, it would be essential to consider the existing infrastructure, the locations of commercial areas, and the movement patterns of

¹The term *Traffic Forecasting* in this dissertation is closer to “교통 예보” in Korean. Translating it as “예측 (prediction)” could lead to the misconception that the focus is solely on resolving the issue of quantitative accuracy.

²The visualization uses data from PEMS-BAY from [Li et al., 2018]. This dataset represents Santa Clara County, near Silicon Valley, also known as the Bay Area. It consists of a total of 24 scenes at 5-minute intervals over three days in 2017, from 4:00 PM to 5:55 PM. The problem is to forecast 12 future scenes (1 hour) based on 12 past scenes (1 hour).

³Platooning is a technology that enables two or more vehicles or trucks to form a convoy and operate autonomously in a coordinated manner, similar to a train. This approach can mitigate phenomena like *traffic waves*, where traffic congestion arises without any clear cause.

urban residents over time. By comprehensively analyzing these factors and the temporal dynamics of traffic flow, it would become possible to forecast traffic patterns effectively.

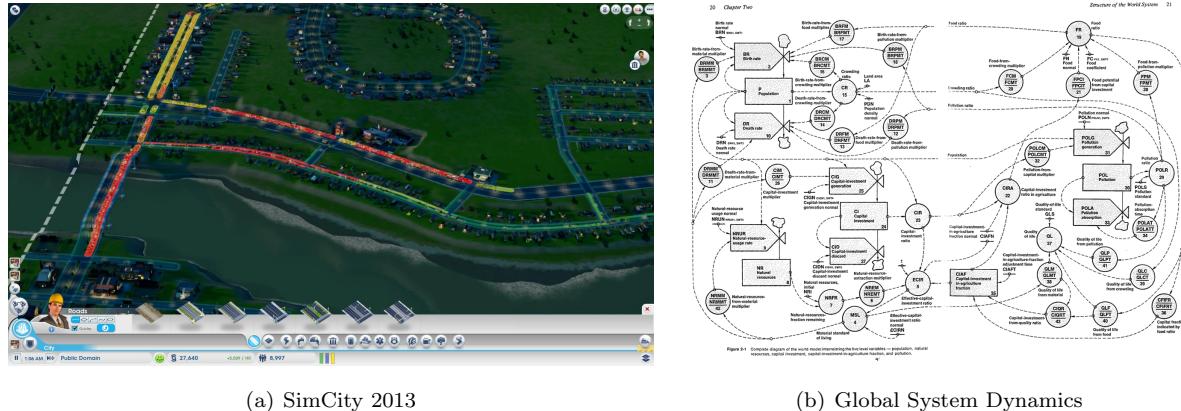


Figure 1.2: (a) Traffic Congestion as Depicted in the SimCity 2013 (b) Forrester's Global System Dynamics Diagram (1971)

During my doctoral research, the experience of playing **SimCity 2013**[contributors, nd], as shown in Figure 1.2[Entertainment, nd], greatly helped me broaden my understanding of cities and the underlying causes of traffic flow. SimCity is a city simulation game first published in 1989 by Will Wright, inspired by Jay Wright Forrester's *Urban Dynamics*[Forrester, 1969]. A priori, phenomena such as traffic congestion, public health issues, crime, and air pollution can be viewed as the comprehensive results of human activities in urban settings. Jay Wright Forrester endeavored to represent the explicit causal relationships between urban variables—such as birth rates, mortality rates, food supply, pollution, capital, and natural resources—through causal inference graphs[Forrester, 1971]. Will Wright adopted this philosophy and built upon it, creating urban simulations based on the principles of agent-based modeling.

However, as times have changed, **SimCity** no longer enjoys the same popularity or resonance it once had. Human behavior is neither predictable nor simplistic, and the diversity in individuals' occupations, preferences, and personalities makes city simulations that yield expected results by constructing purpose-specific buildings less relatable. For instance, in SimCity, buildings are broadly categorized into residential, commercial, and industrial zones. The simulation depicts population movements such as people traveling from residential areas to commercial and industrial zones during early morning hours (5–6 AM), shifting predominantly to commercial zones around noon (12–4 PM), and returning to residential areas in the evening (6–9 PM)⁴. However, these urban mobility patterns vary significantly depending on the city's unique characteristics and its residents' individuality, making them less relatable in many contexts. Additionally, in cities like those in South Korea, where commercial districts remain vibrant even after midnight, SimCity's simulation—based on American urban development models—fails to resonate with or accurately reflect such dynamics.

Before delving into this dissertation, it is important to clarify that it does not address the Traffic Forecasting problem involving entirely novel urban transformations. Furthermore, I fundamentally believe that solving this issue would require an approach grounded not in Jay Forrester's *System Dynamics* but rather in Schopenhauer's philosophy of *Wille*(will)[Schopenhauer, 1833]. However, since this perspective lies outside the mainstream of the CS/AI academic community, it is not discussed in this dis-

⁴Refer to a recorded video of a SimCity day: <https://youtu.be/qqfUsyYwAcM>

sertation. That said, with the advancement of artificial intelligence and robotics, I anticipate a growing shift away from lifestyles centered on repetitive and predictable tasks or labor. Instead, more people are likely to engage in irregular and spontaneous behaviors, further complicating traditional urban mobility models and forecasting challenges.

Returning to the focus of this dissertation, it specifically addresses the **Short-term Traffic Forecasting** problem for freeways, as depicted in Figure 1.1, based on traffic speed data spanning a relatively short period of 2–4 months. Using deep learning techniques, I implemented a model inspired by sequence-to-sequence approaches used in language models, where recurring traffic sensor signal sequences are read to predict the next sequence. The initial problem formulation can be traced back to DCRNN[Li et al., 2018], which is widely regarded as a seminal work in this field. However, the proposed UAGCRN[Han et al., 2023] in this dissertation fundamentally critiques the graph connectivity data definitions in DCRNN. It also provides a broader critique of data-driven graph connectivity learning models, such as GTS[Shang et al., 2021], GraphWaveNet[Wu et al., 2019], and STEP[Shao et al., 2022]. Therefore, the validation of this dissertation’s findings requires a cautious and rigorous approach. Nonetheless, the UAGCRN model is relatively simple, intuitive, and built on a scientific foundation, making it more accessible for verification⁵. While I personally assess the application of the concept of **Human Activity** to deep learning models as still being largely theoretical, this work represents one of the first academic attempts to integrate the humanities concept of human activity within cities into a machine learning framework. This effort underscores a novel interdisciplinary approach, bridging urban studies and AI.

1.2 Challenge

Unlike typical time-series forecasting problems, traffic forecasting requires inferring a sensor’s traffic values by leveraging patterns observed in other sensors. In this process, Graph Convolutional Network (GCN) is normally used where the graph represents the adjacency connectivity between the sensors that can give information Previous studies have explored various spatiotemporal models based on Recurrent Neural Networks (RNN) [Li et al., 2018, Yu et al., 2018, Guo et al., 2019, Zhao et al., 2019] and Transformer [Guo et al., 2021, Shao et al., 2022, Jiang et al., 2023], which have shown effectiveness in time series forecasting while incorporating the spatial adjacency between traffic sensors. However, we argue that there still remain key points for improvement in traffic forecasting domain.

1.2.1 Construction of an Accurate Sensor Adjacency Network

Existing studies [Li et al., 2018, Yu et al., 2018, Zhao et al., 2019] construct a sensor adjacency matrix based on distance-based proximity. However, these studies exhibit a deficiency in providing comprehensive justification for an adjacency matrix construction methodology. For example, the authors of DCRNN defines the sensor adjacency matrix as follows:

$$A_{ij} = \begin{cases} \exp\left(-\frac{\text{dist}(v_i, v_j)^2}{\sigma^2}\right), & \text{if } \text{dist}(v_i, v_j) \leq \kappa, \\ 0, & \text{otherwise.} \end{cases}$$

where A_{ij} represents the edge weight between sensor v_i and sensor v_j , $\text{dist}(v_i, v_j)$ denotes the road network distance from sensor v_i to sensor v_j , σ is the standard deviation of distances, and κ is the threshold. However, for the value of σ , if we simple use as standard deviation of distances, it tends

⁵A review from the CIKM conference at the time is included: <https://smhanlab.com/cikm23-review.txt>

to variate significantly depends on dataset for how much we construct connection between the sensors. While will be mentioned again later chapter, the measured standard deviations of distances are 4.97 miles (METR-LA), 3.93 miles (PEMS-BAY), and 6.92 miles (PEMSD7) respectively. However, it is important to note that the specific σ value can significantly fluctuate depending on the measured distances between sensors and density of their connections on the dataset, which can potentially challenge the previous approach. Plus, the authors of DCRNN used a threshold distance κ of 10 miles to construct the adjacency matrix, while there is no sufficient reasoning for this justification.

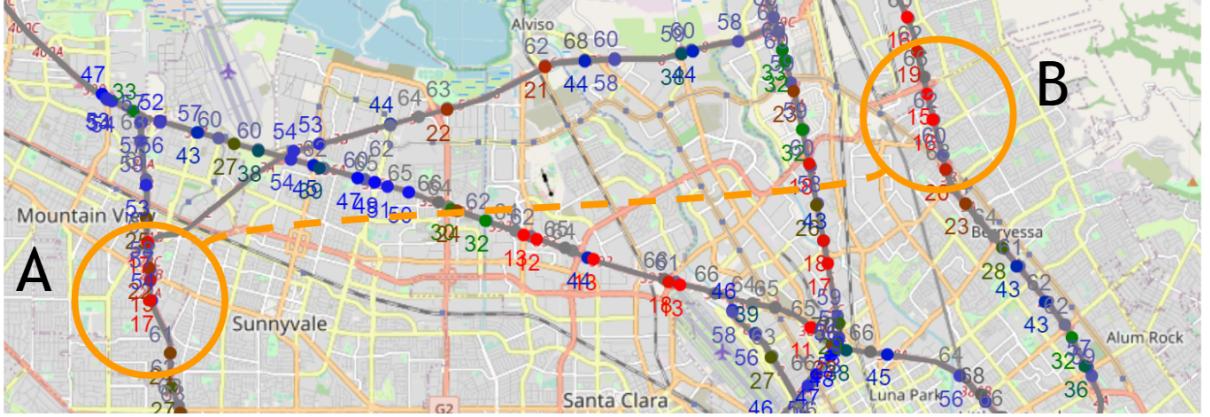


Figure 1.3: Problem of Data-driven Trainable Sensor Adjacency Models.

Recently, several models [Shang et al., 2021, Wu et al., 2019, Guo et al., 2021, Shao et al., 2022] have attempted to propose trainable adjacency matrix by learning from data distribution. For example, GTS [Shang et al., 2021] employs both hard and soft Gumbel-Softmax techniques to construct a trainable sensor adjacency matrix. However, they may generate artificial connections, leading to an inaccurate representation of the sensor network. As illustrated in Figure 1.3, traffic sensors in two geographically distant areas (A, B) may exhibit similar congestion patterns during morning rush hours, where the sensors are not physically connected. If data-driven sensor adjacency is applied, the AI model may infer that these sensors are correlated and mistakenly conclude that they are adjacent.

1.2.2 Addressing Individual Sensor Spatial Heterogeneity

Each traffic sensor is situated within a unique built environment, resulting in diverse congestion patterns. For instance, congestion due to rush hour may occur only in specific lanes or sensors. Even in close locations, different patterns can emerge due to factors such as the number of sensor lanes, entry and exit lanes, and installation positions. Figure 1.4 illustrates that the same speed of 60 miles per hour (mph) may be considered relatively congested at sensors #400723 and #400253, while being a common occurrence at sensor #400514. While previous work such as [Guo et al., 2021] has addressed this issue by leveraging spatial positional encoding, the primary focus has been on handling positional encoding for Transformer models rather than normalizing the patterns of individual sensors.

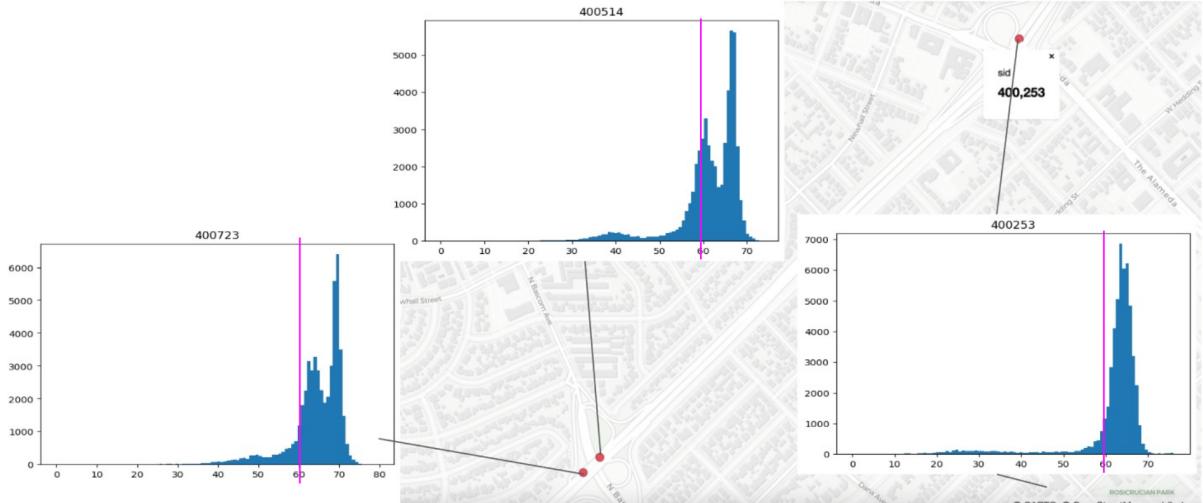


Figure 1.4: Histogram Representing Sensor Heterogeneity in Speed Values.

1.2.3 Integration of Urban Human Activity

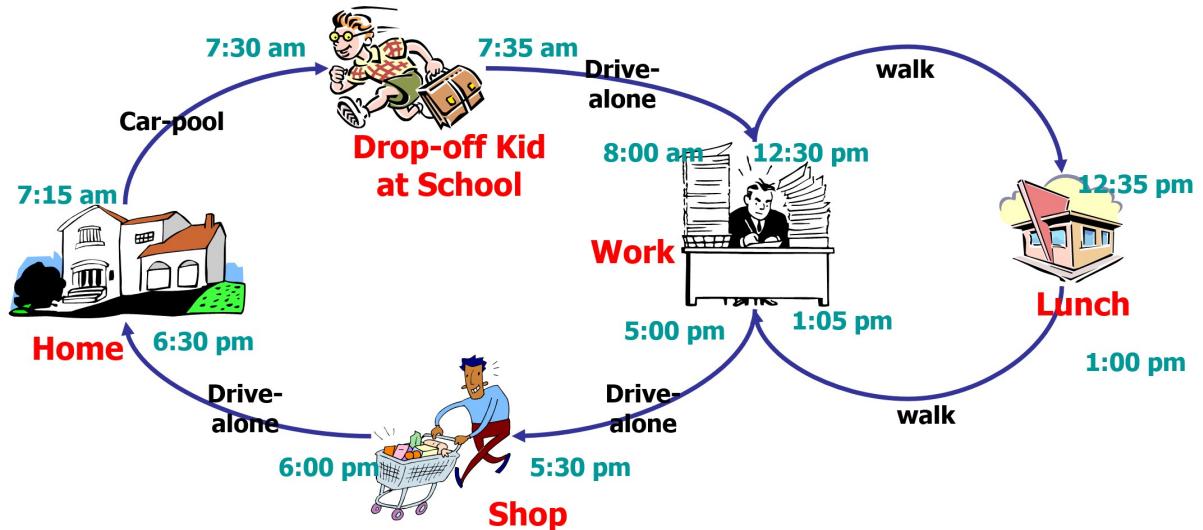


Figure 1.5: Daily Human Activity of a Person

Human activities, such as commuting, significantly influence traffic patterns and can lead to congestion. Figure 1.5[Transportation Research Center (TRC) and Department of Civil and Coastal, 2007] illustrates an example of a person’s travel patterns based on their daily activities. Previous studies have incorporated one-hot encoding of temporal information such as day-of-week and time-of-day (hour, minutes) to capture correlations between time and traffic patterns [Li et al., 2018, Guo et al., 2021, Zheng et al., 2020, Jiang et al., 2023]. For instance, GMAN [Zheng et al., 2020] represents a day with T time steps⁶ and encodes the day-of-week and time-of-day for each time step using one-hot vectors in \mathbb{R}^7 and \mathbb{R}^T , respectively. These encodings are then concatenated into a vector in \mathbb{R}^{T+7} as illustrated on Figure 1.6.

While temporal information offers valuable insights into human activities, it does not inherently establish direct causality, as traffic patterns are driven by human actions. Moreover, conventional temporal

⁶e.g. $T = 288$ when a day is divided into 5-minute intervals.

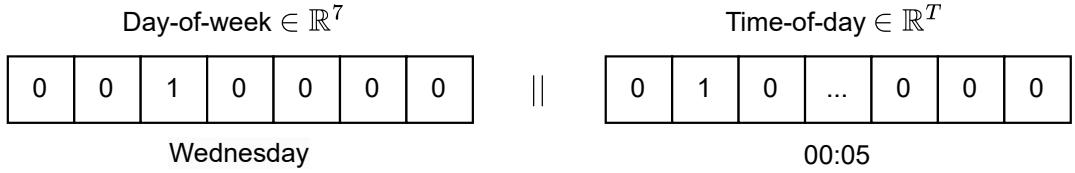


Figure 1.6: Temporal Encodings leveraged in GMAN

encodings, represented by discrete values, often struggle to capture the complexity of continuous spatio-temporal urban traffic congestion scenarios. Thus, it becomes essential to model the correlation between traffic patterns and urban human activities, moving beyond a sole reliance on temporal information.

1.3 Contribution

This dissertation presents a novel framework to tackle the challenges of generating realistic vehicle travel trajectories, aimed at enhancing the construction of sensor connectivity information. The proposed solution leverages the A* algorithm to create travel trajectories and construct a sensor adjacency matrix, which is subsequently integrated into graph-convolutional spatiotemporal models.

To address the spatial heterogeneity of sensor networks, a one-hot-based sensor encoding is employed, tailored to each sensor's unique characteristics. This approach ensures adaptability to diverse sensor environments. Additionally, to capture the relationship between human activity and traffic patterns, diachronic urban travel activity frequencies—estimated at the target forecasting timestamp—are incorporated, utilizing data derived from the National Household Travel Survey [U.S. Department of Transportation, 2017].

The framework includes two spatiotemporal deep learning architectures: UAGCRN and UAGC-Transformer. These models effectively integrate the constructed graph into graph-convolutional recurrent neural networks and graph-convolutional transformers, respectively. While both RNN-based and Transformer-based temporal models are investigated, the results demonstrate that the RNN-based model is sufficient for tackling the time-series traffic forecasting problem.

The proposed UAGCRN achieves state-of-the-art performance on standard traffic datasets, outperforming existing baselines. Furthermore, the effectiveness of the constructed graph is validated by its positive impact on other spatiotemporal models. The scalability of the sensor and activity embedding approach is also demonstrated by its successful application to purely temporal models, such as LSTM [Sutskever et al., 2014] and Transformer [Vaswani, 2017].

To foster further research and development in this domain, the code, dataset, and experiment logs from this study are made publicly available.⁷

⁷<https://github.com/SuminHan/Traffic-UAGCRNTF>

Chapter 2. Literature Review

2.1 Spatiotemporal Deep Learning Traffic Forecasting Models

2.1.1 Proposal of the Traffic Forecasting Problem

In the deep learning community, two key studies that initially proposed the traffic forecasting problem are DCRNN (ICLR-18)[Li et al., 2018] and STGCN (IJCAI-18)[Yu et al., 2018]. DCRNN and STGCN learn spatiotemporal patterns based on traffic speeds measured from traffic sensors and propose incorporating data from non-Euclidean spaces in graph form¹ into time-series models.

The Spatio-Temporal Graph Convolutional Network (STGCN)[Yu et al., 2018] effectively processes spatiotemporal data on graphs by efficiently calculating high-order terms of the graph Laplacian matrix using a Chebyshev polynomial approximation and uses a 1D-CNN to learn temporal changes. The Chebyshev polynomial approximates spectral filters on the graph, allowing for filtering that reflects the relationships between nodes and enables learning interactions between distant nodes. Through this, STGCN implements a graph convolution operation that considers both spatiotemporal features, allowing each node to efficiently learn its surrounding information without needing to calculate eigenvectors.

The Diffusion Convolutional Recurrent Neural Network (DCRNN)[Li et al., 2018] is designed by integrating Diffusion Convolution, which models the diffusion process on a graph, with an RNN structure to solve spatiotemporal forecasting problems like traffic forecasting. Diffusion Convolution uses a bidirectional random walk diffusion process to capture dependencies between nodes on the graph, where each node propagates information from itself and its neighboring nodes. This approach allows DCRNN to simultaneously learn spatial relationships in the traffic network and temporal dependencies through the RNN structure. As a result, DCRNN effectively combines the information from complex graph structures and time-series patterns to predict future states based on spatiotemporal data.

Between the two models (DCRNN and STGCN), I find DCRNN’s problem formulation and model structure to be more intuitive, making it a foundational approach for addressing the traffic forecasting problem in my thesis [Han et al., 2023]. The Seq2Seq traffic forecasting problem defined by DCRNN is equivalent to finding the optimal function f that performs the following:

$$[X_{t-P+1}, \dots, X_t; \mathcal{G}] \xrightarrow{f} [X_{t+1}, \dots, X_{t+Q}] \quad (2.1)$$

Here, $X_t \in \mathbb{R}^{N \times C}$ represents the values from traffic sensors, with N sensors and C traffic channels (such as speed, traffic volume, etc.). For the graph, $\mathcal{G} = (\mathcal{V}, \mathcal{E}, \mathbf{A})$ generally defines sensors as nodes, with $|\mathcal{V}| = N$, and \mathcal{E} represents edges, which indicate adjacency information between sensors. The adjacency matrix $\mathbf{A} \in \mathbb{R}^{N \times N}$ reflects the information in \mathcal{E} and is used later in the Graph Convolutional Network (GCN).

2.1.2 Traffic Datasets

International traffic forecasting research currently primarily utilizes U.S. data. Figure 2.1 shows the METR-LA and PEMS-BAY datasets used in DCRNN, and the PeMSD7 (M, L) data used in STGCN.

¹While 2D-Grid represented by x and y coordinates is referred to as Euclidean space, graph structures can be considered as non-Euclidean space.

These datasets are based on traffic information from Los Angeles (LA) and Santa Clara, near Silicon Valley, in California, USA. STGCN uses LA data divided into PeMSD7(M) and PeMSD7(L), with PeMSD7(M) covering a relatively smaller area and PeMSD7(L) a larger one. Notably, the PeMSD7(M, L) data deals with urban data that includes downtown areas like Hollywood, unlike METR-LA.

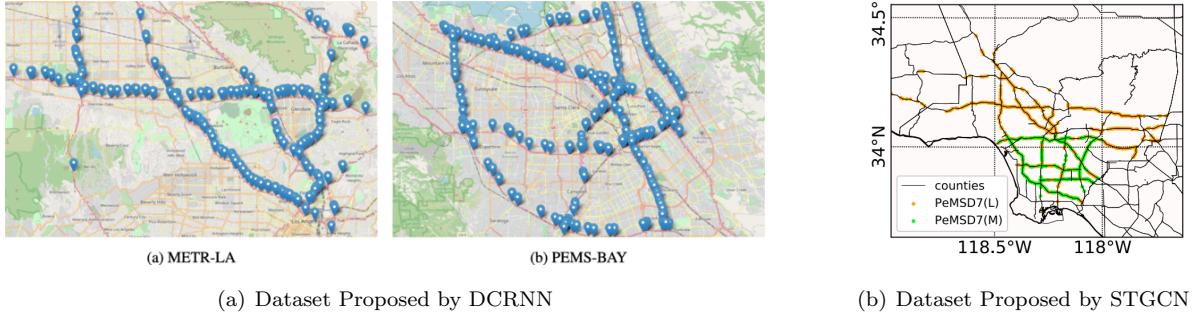


Figure 2.1: (a) METR-LA and PEMS-BAY from DCRNN (b) PeMSD7 (M, L) from STGCN

2.1.3 Diffusion Convolutional Recurrent Neural Network

DCRNN[Li et al., 2018] proposes the Diffusion Convolutional Recurrent Neural Network to solve the traffic forecasting problem as Equation 2.2.

$$\mathbf{X}_{:,p} \star_{\mathcal{G}} f_{\theta} = \sum_{k=0}^{K-1} \left(\theta_{k,1} (D_O^{-1} W)^k + \theta_{k,2} (D_I^{-1} W^{\top})^k \right) \mathbf{X}_{:,p} \quad \text{for } p \in \{1, \dots, P\} \quad (2.2)$$

DCGRU is a model that applies graph convolution ($\star_{\mathcal{G}}$)² to GRU[Chung et al., 2014]. Similar to ConvLSTM[Shi et al., 2015], it performs graph convolution within the internal r , u , and C units that constitute the GRU cell. The structure is further detailed in Equation 2.3.

$$\begin{aligned} r^{(t)} &= \sigma \left(\Theta_r \star_{\mathcal{G}} \left[X^{(t)}, H^{(t-1)} \right] + b_r \right), \\ u^{(t)} &= \sigma \left(\Theta_u \star_{\mathcal{G}} \left[X^{(t)}, H^{(t-1)} \right] + b_u \right), \\ C^{(t)} &= \tanh \left(\Theta_C \star_{\mathcal{G}} \left[X^{(t)}, (r^{(t)} \odot H^{(t-1)}) \right] + b_c \right), \\ H^{(t)} &= u^{(t)} \odot H^{(t-1)} + (1 - u^{(t)}) \odot C^{(t)}. \end{aligned} \quad (2.3)$$

Then, this approach is used in Sequence-to-Sequence (Seq2Seq) modeling to predict the short-term future. Typically, traffic forecasting utilizes data in 5-minute intervals, where the model reads the past 12 time steps (equivalent to 1 hour of patterns) and predicts the next 12 time steps corresponding to

²In general, the operation in one layer of a GCN is expressed as follows:

$$H^{(l+1)} = \sigma \left(\tilde{D}^{-\frac{1}{2}} \tilde{\mathbf{A}} \tilde{D}^{-\frac{1}{2}} H^{(l)} W^{(l)} \right)$$

Each term is defined as follows: $H^{(l)}$ is the node feature matrix at the l -th layer, $W^{(l)}$ is the learnable weight matrix of the l -th layer, σ is the activation function, $\tilde{\mathbf{A}} = \mathbf{A} + I$ is the adjacency matrix with self-loops added, and \tilde{D} is the diagonal normalization matrix of $\tilde{\mathbf{A}}$ ($\tilde{D}_{ii} = \sum_j \tilde{\mathbf{A}}_{ij}$). This formula represents the process of aggregating node features on the graph from neighboring nodes and applying a non-linear transformation at each layer through learnable weights and an activation function.

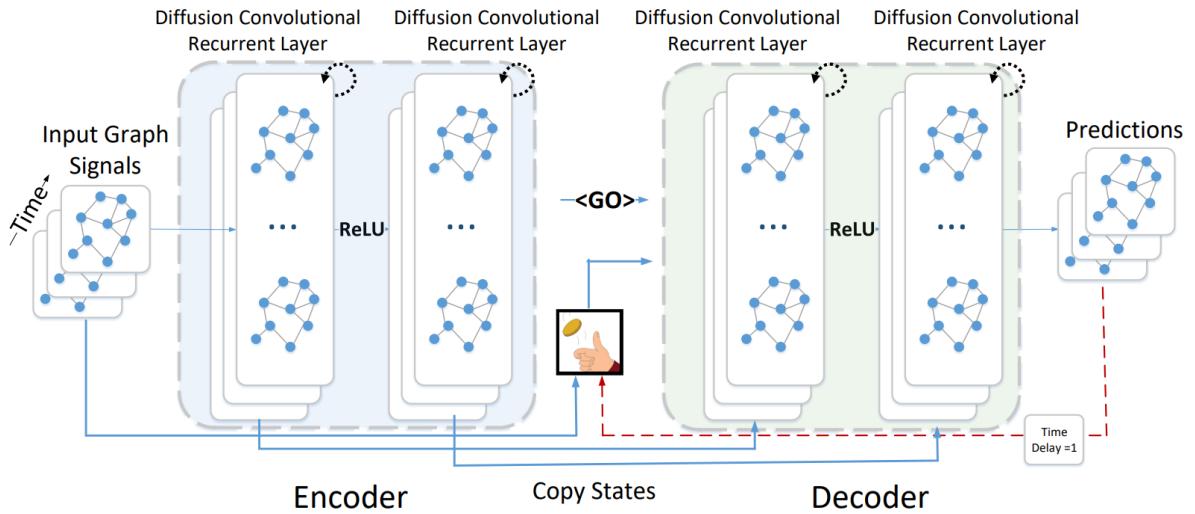


Figure 2.2: Sequence to Sequence Encoder-Decoder of DCRNN

the following hour. Figure 2.2 shows the Seq2Seq structure of DCRNN, which uses an encoder-decoder structure to make these forecastings. Similar to FC-LSTM[Seo et al., 2018a], this structure passes the state of the encoder to the decoder, and the forecasting process begins by providing a “Go Token” (usually a zero value).

2.2 Advanced Spatiotemporal Traffic Forecasting Models

Traffic forecasting research has advanced through various methodological approaches. Figure 2.3 visually shows the major development trends in traffic forecasting models, from DCRNN[Li et al., 2018] and STGCN[Yu et al., 2018] to the UAGCRNN[Han et al., 2023] proposed in this study.

First, FC-LSTM[Seo et al., 2018a], a Seq2Seq structure commonly used in language modeling, was one of the initial models introduced for traffic forecasting, providing a foundation for time-series data forecasting. Subsequently, DCRNN[Li et al., 2018] and STGCN[Yu et al., 2018] evolved from FC-LSTM by integrating the temporal patterns of time-series data with the spatial relationships of road networks. DCRNN was the first model to effectively reflect the spatiotemporal characteristics of traffic forecasting by integrating recurrent neural networks (RNN) with graph structures, particularly contributing by modeling interactions between adjacent roads through graph neural networks. STGCN, which combined graph convolution and CNN structures, improved computational efficiency but was considered less interpretable than DCRNN.

Later, GeoMAN[Liang et al., 2018] and ASTGCN[Guo et al., 2019] introduced attention mechanisms, adding the capability to emphasize important spatiotemporal factors in traffic forecasting. The attention technique helps capture factors that impact traffic congestion by emphasizing certain times or regions within the data. Graph WaveNet[Wu et al., 2019] applied the WaveNet[Oord, 2016] structure, known for its strength in signal processing, to graph-based traffic forecasting, contributing to more precise temporal pattern forecasting. These models achieved significant technical advancements in traffic forecasting but were somewhat limited by their tendency to simply adopt popular structures from other AI research.

The Transformer[Vaswani, 2017] structure has had a substantial impact on AI research since its introduction in 2017, showing high potential for application in traffic forecasting as well. GMAN[Zheng et al., 2020], which applies this structure to traffic forecasting, uses the Transformer's strength in time-series processing to address spatiotemporal information together. However, there is still debate over whether the Transformer's advantages are fully realized in traffic forecasting, with some critiques suggesting a need for specific approaches that optimize traffic data characteristics rather than following structural trends.

More advanced models include GTS[Shang et al., 2021] and ST-ODE[Fang et al., 2021]. GTS introduces the Gumbel Softmax[Jang et al., 2016, Maddison et al., 2016] technique to infer the adjacency relationships needed for traffic forecasting, eliminating the need for pre-knowledge of sensor connection information. ST-ODE applies ordinary differential equations (ODEs) to analyze traffic data as a continuous time flow, enhancing forecasting accuracy over time. Additionally, STEP[Shao et al., 2022] and PDFormer[Jiang et al., 2023] propose an approach to improve forecasting performance by dividing traffic data into small patches and training them within the Transformer structure. However, these models have also faced criticism for a focus on minor performance improvements through the latest technology without fundamental consideration of model persuasiveness or reliability.

Since 2019, deep learning-based models have been actively published in traffic-related journals. Notable examples include T-GCN[Zhao et al., 2019] and Traffic Transformer[Cai et al., 2020], which aim not only to improve performance but also to address practical issues in traffic forecasting. Journal papers generally focus on real-world problems, distinguishing them from conference research, which predominantly presents pure AI research. Journals like Transactions on Intelligent Transportation Systems (TITS), Transactions on GIS, and Transportation Research Part B and C focus more on models geared towards practical problem-solving in areas such as traffic forecasting.

This study proposes UAGCRN[Han et al., 2023] to overcome the limitations of existing research, moving away from trend-following structural adoption and aiming for model design that deeply understands and reflects the characteristics of traffic data. In particular, this model aims to account for non-standard variations beyond repetitive patterns in traffic forecasting, addressing human mobility and lifestyle patterns to contribute to practical problem-solving.

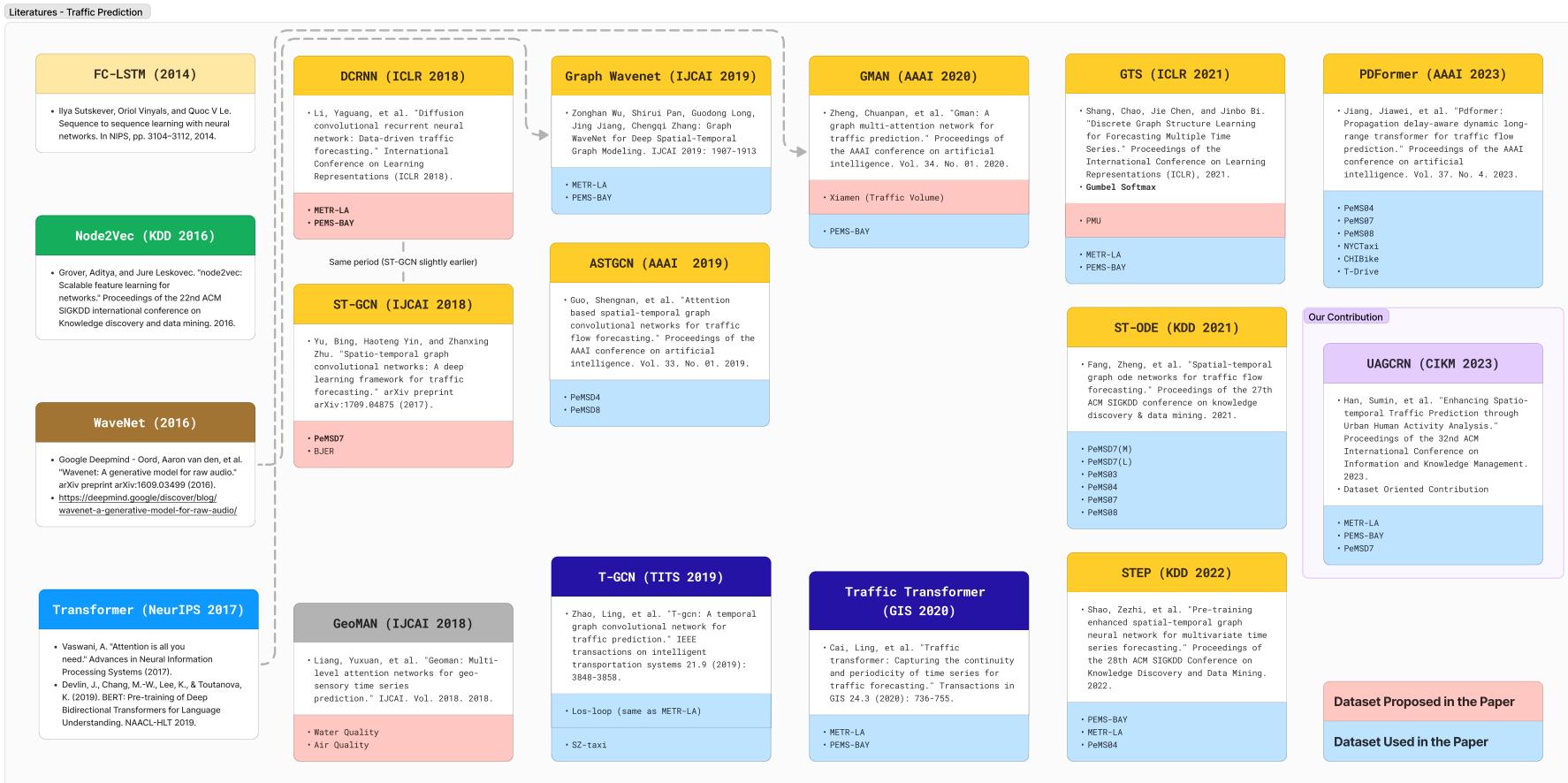


Figure 2.3: Mainstream Traffic Forecasting Models, with our model (UAGCRN) highlighted in purple.

Criticism

Existing studies have primarily adopted or modified model structures with the goal of improving the performance of traffic forecasting models by incorporating the latest AI techniques. Conferences focused on pure AI, such as NeurIPS, typically emphasize neural network techniques aimed at enhancing pattern recognition or replacing human sensory functions, with a strong focus on performance improvement. However, this approach diverges from the goal of designing AI that understands and assists humans. Such studies often lack exploration of human actions and their causes, or fail to deeply consider the specific requirements of a given domain. This contrast highlights the limitations of trend-driven research in traffic forecasting, as opposed to the direction sought by HCI-AI research.

Most methods introduced in previous studies have been driven more by trends in AI research than by a scientific or domain-specific approach. For instance, GraphWaveNet was inspired by the release of WaveNet, while the popularity of the Transformer model led to the development of GMAN in traffic forecasting. Although leveraging such trends can be valuable in some contexts, adopting them without a deep understanding of traffic systems or human behavior risks leading to inconsistent research directions. This lack of coherence, especially in a field like traffic forecasting that requires practical and reliable solutions, can undermine the long-term value and impact of the research.

This trend often results in excessive focus on minor decimal-level performance improvements, and in traffic forecasting, such minute enhancements are frequently accepted as significant contributions. While this approach may be relevant in the early stages of research, questions arise as to whether models developed for short-term performance improvement offer lasting usefulness. On the other hand, models like DCRNN[Li et al., 2018], which reflect the spatiotemporal patterns of traffic data effectively beyond their technical achievements, remain valuable, indicating the need for a scientific and fundamental consideration in model development.

This dissertation aims not only for performance improvement but also for a deep understanding and integration of human actions and their causes in traffic forecasting. Beyond merely learning data patterns, traffic forecasting should reflect rapidly changing urban environments and irregular human behaviors. This study seeks to enhance practical problem-solving capabilities by incorporating human-centered thinking into traffic forecasting models.

LibCity

LibCity[Wang et al., 2021] team³ has organized the code for traffic forecasting models and neatly arranged datasets, as shown in Figure 2.4. As of 2024, more than 56 traffic forecasting models and approximately 53 traffic datasets have been implemented and made publicly available. This collection includes not only traffic forecasting but also code for various problems such as transportation demand forecasting and traffic OD (Origin-Destination) flow forecasting.

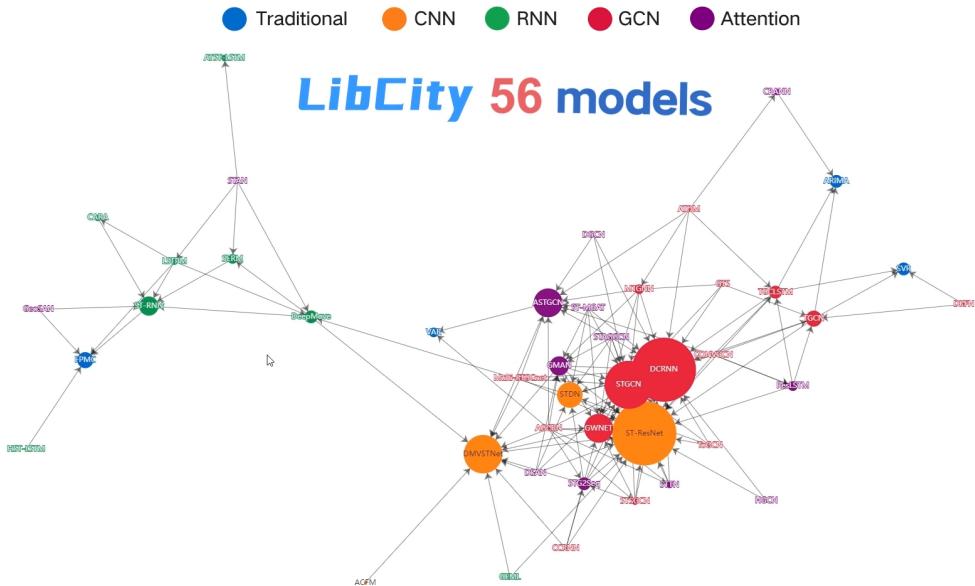


Figure 2.4: Implemented Spatiotemporal models in LibCity

³<https://libcity.ai/>

2.3 Human-Centered Understanding of Urban Mobility

This section introduces concepts for modeling improvements aimed at enhancing traffic forecasting through a human-centered understanding. Unlike conventional deep learning approaches that primarily model traffic flow systematically, this approach seeks to understand traffic phenomena from a human-centered perspective.

2.3.1 Traffic Wave Theory

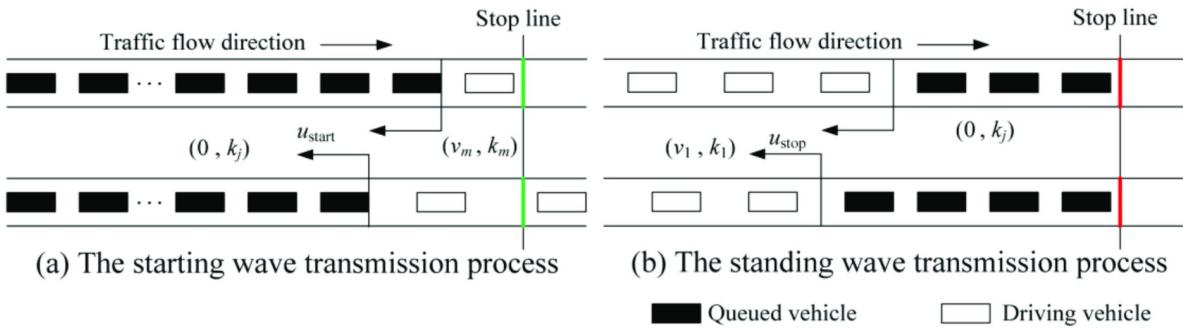


Figure 2.5: Bidirectional Traffic Waves

Traffic Wave Theory, as illustrated in Figure 2.5[Wang et al., 2022], explains the phenomenon where traffic congestion, particularly on highways, spreads like a wave, also referred to as “Traffic Snakes” or “Traffic Shocks”. This theory demonstrates that traffic flow can propagate forwards and backwards, similar to physical waves, and is useful for explaining how human factors such as drivers’ reaction times and distance control impact traffic congestion. By analyzing the patterns of congestion waves generated by drivers, Traffic Wave Theory provides a foundation for traffic forecasting models to incorporate human behavioral patterns and psychological factors.

The Dual-Walk Graph Convolution proposed in this dissertation (Section 3.2.4) is designed based on Traffic Wave Theory, allowing simulation of bidirectional congestion wave propagation through 1-diffusion while excluding DCRNN’s k-diffusion. In this process, Traffic Wave Theory enables more realistic and interpretable traffic forecastings by reflecting the specific patterns of congestion propagation beyond mere temporal data forecasting. By considering that traffic congestion forms and dissipates based on factors such as drivers’ reaction speeds and distance-maintenance habits, this approach helps the model adopt a human-centered approach in congestion scenarios.

While Traffic Wave Theory is particularly effective in explaining traffic flows on highways, its applicability is limited in complex areas such as urban intersections or regions with intricate signal systems. In urban environments, where various signals and physical obstacles disrupt the uniformity of traffic flow, the propagation of traffic patterns does not resemble wave-like behavior. Consequently, this dissertation focuses on applying Traffic Wave Theory to freeway-based forecasting, recognizing that additional theoretical frameworks are necessary to model more complex environments like urban intersections.

This human-centered approach not only aims to predict physical flow but also to improve the reliability of traffic forecasting by reflecting how humans react and behave within traffic flows. If the model can incorporate drivers’ reactions in congestion scenarios through Traffic Wave Theory, it will provide a valuable foundation for solving practical issues, such as traffic congestion management and

accident prevention.

2.3.2 Traffic Speed-Volume Relationship

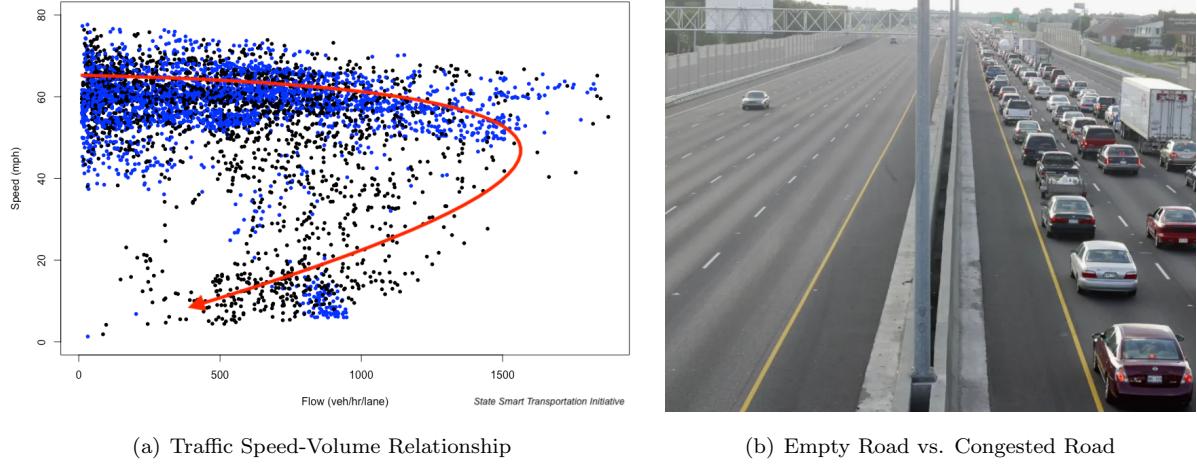


Figure 2.6: (a) Nonlinear Relationship Between Traffic Speed and Traffic Volume (b) Empty Road vs. Congested Road

Traffic speed and traffic volume have a nonlinear relationship, which serves as an essential conceptual basis for traffic forecasting modeling. For instance, as shown in Figure 2.6 (a)⁴, traffic volume increases as speed increases until it reaches around 50 mph, after which the traffic volume no longer rises. This can be easily understood in the context of highway conditions. When there are many vehicles on the road, an increase in speed allows more vehicles to pass through. However, when the road is almost empty, increasing speed does not significantly increase the actual traffic volume, demonstrating a nonlinear characteristic. This relationship can be intuitively understood by comparing the empty and congested roads shown in Figure 2.6 (b)⁵.

This nonlinear relationship between traffic speed and traffic volume is one of the key reasons deep learning models are effective at learning complex patterns. Nonlinear functions can be learned through the activation functions in deep learning; by stacking multiple layers of activation functions like ReLU, these models can effectively capture nonlinear relationships. Based on this concept, the UAGCRN model in this study uses a 2-stacked fully connected layer in the initial layer to learn the nonlinear relationship between traffic speed and volume. This initial layer provides sufficient nonlinearity to specifically reflect the relationship between speed and volume, enabling traffic volume forecasting that aligns with the characteristics of the data.

The use of a 2-stacked fully connected layer in the initial layer of UAGCRN was inspired by the structure of the GMAN model, but it goes beyond simply borrowing the structure. This study designed this initial layer with the understanding that this nonlinear relationship is an essential feature of traffic speed and volume forecasting. Experiments confirmed that learning this nonlinear relationship leads to more accurate and interpretable results in real traffic volume forecastings. In particular, it enables real-time adaptation to changes in the relationship between speed and volume, allowing for more flexible performance across various traffic scenarios than traditional linear models.

⁴Figure from With traffic down, Oregon DOT can move more vehicles at twice the speed, By Chris McCahill

⁵Mario Villafuerte/Getty Images – Why do traffic jams sometimes form for no reason?

By effectively learning the nonlinear relationship between traffic speed and volume, the UAGCRN model reflects traffic flow variability and improves traffic forecasting performance in both congested and high-speed sections. This design, which incorporates the nonlinear relationship, is a foundational element that enhances interpretability and reliability in traffic volume forecasting, providing a methodological basis for accurately capturing irregular patterns in real-world traffic flow.

2.3.3 Activity-Based Household Travel Surveys

Activity-based survey data, such as the U.S. National Household Travel Survey[U.S. Department of Transportation, 2017] (Figure 2.7), Korea's Household Travel Survey, and Floating Population data, are critical for understanding human movement patterns and purposes, providing essential information for traffic forecasting. These datasets classify origins and destinations as Home, Work, and Else, or analyze destinations by time of day, allowing for visualization of movement patterns at specific times and locations. For instance, Korea's Floating Population data categorizes the origin and destination of regional populations on an hourly basis, helping to accurately capture travel purposes and patterns.

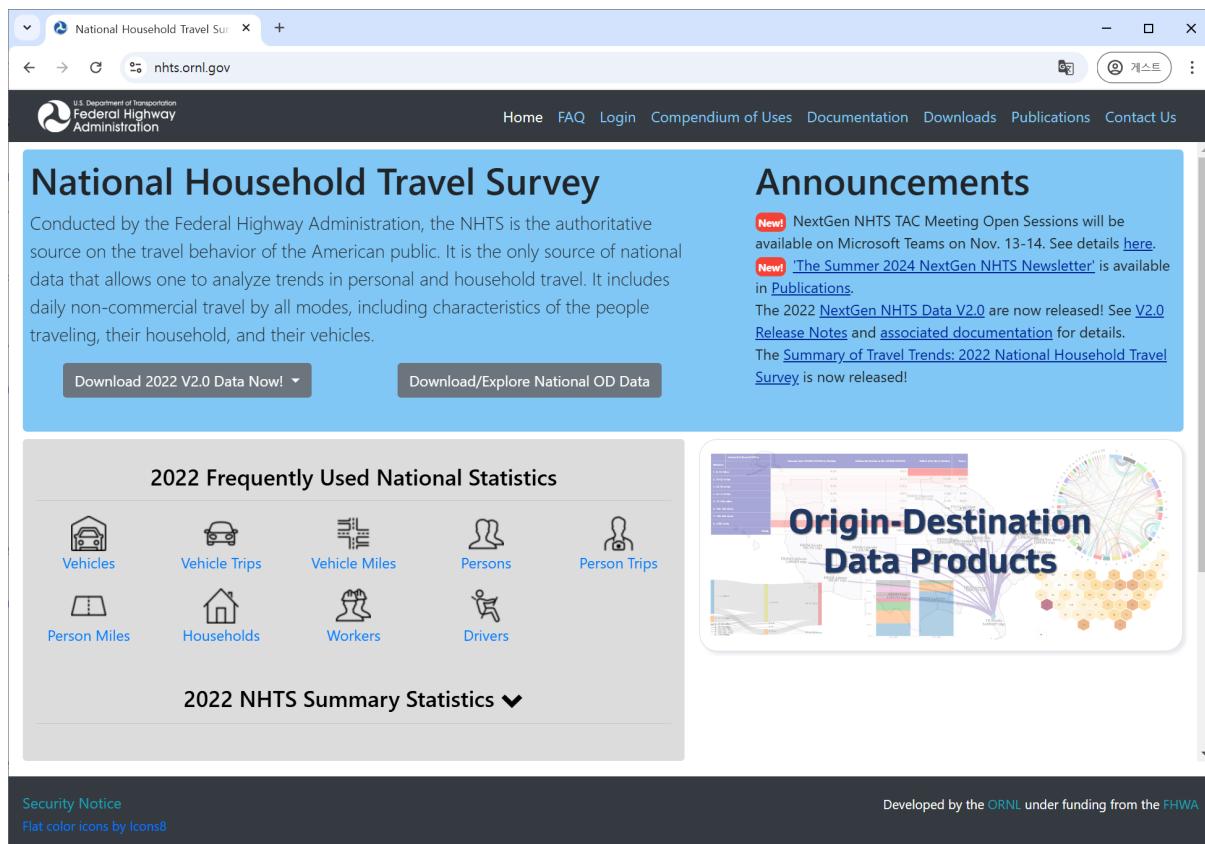


Figure 2.7: National Household Travel Survey Webpage

Activity-based data from the U.S. and Korea is useful for comparing movement patterns between countries and analyzing how regional characteristics or cultural differences influence travel behavior. For example, Korea's Floating Population data displays real-time movement patterns, which can reveal congested areas and overcrowded zones during specific time periods. Such comparisons play an important role in enhancing model flexibility and developing models that can be applied to various urban environments.

Figures 2.8 and 2.9 provide examples of Floating Population data and a graph visualizing movement purposes by time and day of the week. These help predict the times and locations where traffic congestion is likely to occur. In particular, Figure 2.9 clearly shows the differences in weekday commute times versus weekend activity patterns, assisting traffic forecasting models in reflecting different travel characteristics by time of day.

ym	weekday	hour	orient	destination	gender	age	type	duration	val	OD	WK	HH	orient_str	destination_str	
0	202112	일	23	1101053	1101053	F	15	WH	10	12.12	1101053-1101053	6	167	1101053	1101053
1	202112	일	23	1101053	1101053	F	30	EE	10	2.39	1101053-1101053	6	167	1101053	1101053
2	202112	일	23	1101053	1101053	F	30	WH	10	2.35	1101053-1101053	6	167	1101053	1101053
3	202112	일	23	1101053	1101053	F	30	EH	10	2.37	1101053-1101053	6	167	1101053	1101053
4	202112	일	23	1101053	1101053	F	30	HW	10	2.35	1101053-1101053	6	167	1101053	1101053
...	
316754	202112	일	23	39000	1125074	F	55	EH	80	3.25	39000-1125074	6	167	39000	1125074
316755	202112	일	23	39000	1125074	M	10	EH	30	6.55	39000-1125074	6	167	39000	1125074
316756	202112	일	23	39000	1125074	M	30	EH	70	2.36	39000-1125074	6	167	39000	1125074
316757	202112	일	23	39000	1125074	M	45	EH	50	3.16	39000-1125074	6	167	39000	1125074
316758	202112	일	23	39000	1125074	M	50	EH	60	6.54	39000-1125074	6	167	39000	1125074

Figure 2.8: Sample of Seoul Floating Population Movement Data

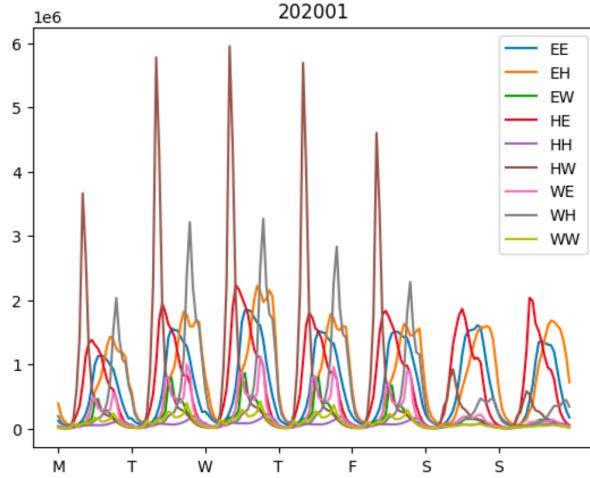


Figure 2.9: Graph of Floating Population Movement Data in Seoul by Day and Hour

However, survey and floating population data have certain limitations. Surveys rely on respondents' subjective assessments, which makes it challenging to perfectly capture actual movement patterns, and they also have limitations in collecting real-time data. Floating population data, likewise, only reflects movement in specific regions or times, which can result in temporal or spatial constraints. To overcome these limitations, complementary methods are needed, such as combining activity-based data with real-time data or introducing more detailed analysis techniques to improve the accuracy of movement pattern analysis.

Such an activity-based approach holds value beyond merely collecting movement data. By identifying movement patterns based on specific times and purposes, it is possible to provide tailored traffic services to anticipated congestion areas or to plan public transportation deployment according to travel purposes. This study aims to incorporate activity-based data into traffic forecasting models, developing a predictive model that considers movement characteristics by time and destination, ultimately contributing to solving real-world problems.

2.3.4 Origin-Destination Based Movement Behavior

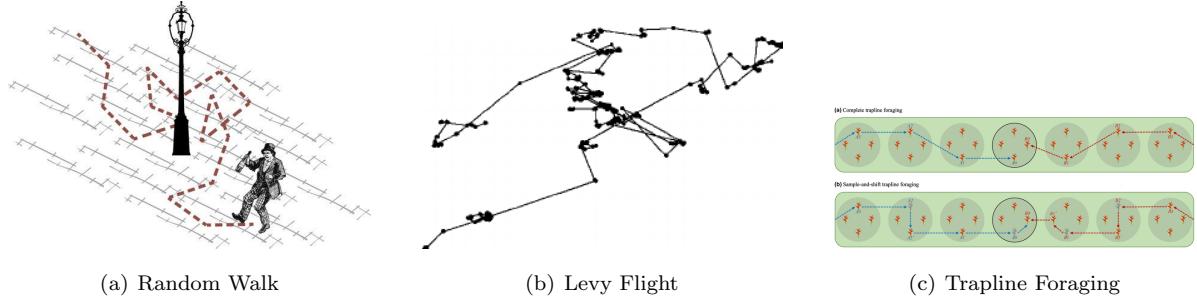


Figure 2.10: (a) Random Walk (b) Levy Flight (c) Trapline Foraging

There are various models to simulate human movement patterns, each reflecting specific characteristics of movement behavior. For example, the Random Walk model represents a simple pattern of random movement (Figure 2.10 (a)), while the Levy Flight model involves occasional long-distance movements amid random movement (Figure 2.10 (b)). The Trapline Foraging model simulates the pattern of animals following a regular route to search for food, allowing for the modeling of repetitive yet efficient paths (Figure 2.10 (c)).

These models are useful for understanding movement patterns, but humans differ from animals in that they tend to prefer the shortest path to a destination. For instance, humans typically move toward a specific destination rather than engaging in random movement without purpose and tend to optimize routes based on factors such as traffic conditions or efficiency. Moreover, humans engage in a wider range of activities and follow diverse routing plans, making it challenging to apply straightforward walking models. Therefore, simple models like the Random Walk or Trapline Foraging models are insufficient to fully explain human movement patterns. However, ethical concerns may arise in directly studying human movement patterns, so some research approaches gain insights into human movement by studying animal movement patterns. For example, previous studies using movement data from animals, such as cats, serve as one example of this approach [Bischof et al., 2022].

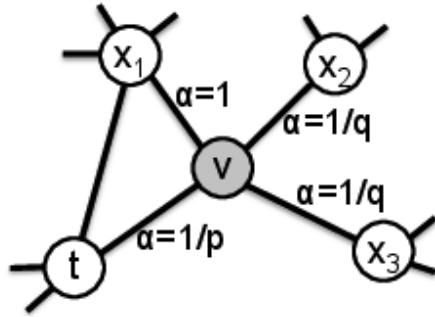


Figure 2.11: Node2Vec Illustration

The Node2Vec model, a random walk-based approach, is designed to capture relationships between nodes in a graph by calculating the probability of returning to a specific node (p) and the probability of jumping to a new node (q), enabling the exploration of various paths within the graph (Figure 2.11). While this is effective for graph-based network analysis where random exploration is crucial, it differs significantly from human movement patterns, which are more destination-oriented. Human movement

typically focuses on efficiently reaching specific destinations, necessitating methods that account for destination-driven behavior rather than simple random walks. This limitation is also highlighted in Road2Vec[Wang et al., 2020], where the authors note that conventional approaches relying on random walks fail to accurately capture the movement behaviors of mobile road users, who often prioritize shortest paths when traveling from sources to destinations.

In this dissertation, urban areas are divided into grids, and the A* algorithm is employed to simulate probable travel paths. To introduce variability, a small probabilistic variation is applied to account for the likelihood of using freeways. Additionally, the co-occurrence probabilities of sensors are used to construct a sensor adjacency matrix, which is then utilized in a graph convolutional recurrent neural network. However, this random path generation could be enhanced by incorporating a deeper understanding of human activities, leveraging factors such as land use, transportation services, and points of interest (POIs).

Chapter 3. Human-Activity-based Traffic Forecasting Models

3.1 Problem Formulation

We begin by formally defining the problem of spatiotemporal traffic prediction. We introduce a sensor adjacency graph $\mathcal{G} = (V, E, \mathbf{A})$, where V represents the set of sensors, E denotes the set of edges representing sensor adjacency, and \mathbf{A} represents the adjacency matrix. Hence, $N = |V|$ signifies the number of traffic sensors in the graph. At each time step t , the traffic values of the N sensors are represented by $X_t \in \mathbb{R}^N$. Additionally, we consider the frequency of urban human activity at time step t , denoted as $H_t \in \mathbb{R}^{K_H}$, where K_H indicates the number of categories for human activity.

Our problem is to learn a function f that predicts the next Q timesteps of traffic values, given a historical sequence of P timesteps of traffic values and $P + Q$ timesteps of estimated human activity frequencies $H_{t-P+1}, \dots, t+Q$.

$$[(X_{t-P+1}, H_{t-P+1}), \dots, (X_t, H_t); \mathcal{G}] \xrightarrow{f} [X_{t+1}, \dots, X_{t+Q}] \quad (3.1)$$

3.2 Methodology

In this section, we present the methodology employed to address the traffic prediction problem. Our approach is composed of three key contributions: refined proximity graph construction, spatial sensor embedding, and urban activity embedding. These contributions form the foundation for our model architectures, namely UA-GCRN and UA-GCTransformer, which incorporate the Graph Convolutional Recurrent Network (GCRN) and attention-based Graph Transformer, respectively, as illustrated in Fig.3.4 and 3.5. The term **UA** denotes the combination of sensor embedding (SE) and activity embedding (AE) added to the input of the encoder and decoder, to distinguish our approach.

3.2.1 Graph Construction

Travel path generation

We partition the region in which traffic sensors are located into a grid of size $N_H^{(\text{Grid})} \times N_W^{(\text{Grid})}$. To generate plausible paths for each pair of grid regions, we employ the A* algorithm [Hart et al., 1968], resulting in a set of paths, $\mathcal{M}^{(\text{Gen})}$. The A* algorithm efficiently explores the search space by combining uniform cost search and greedy best-first search. It selects the edge with the minimum cost as it progresses. In the A* algorithm, the cost of the next step movement is determined by the distance of the road, enabling the identification of the shortest path. By applying a coefficient to the cost of freeways during path generation, less than 1, we can obtain multiple paths with varying levels of freeway usage (see Figure 3.1). An exemplary stacked visualization of the generated travel paths and sensor appearances along a path is presented in Figure 3.2.

Sensor Adjacency Matrix Construction

Firstly, we define the distance between two sensors, v_i and v_j , taking into account the road direction that can be traveled by car. Here, $i, j \in 1, \dots, N$ represent the sensor indices. In our research, traffic

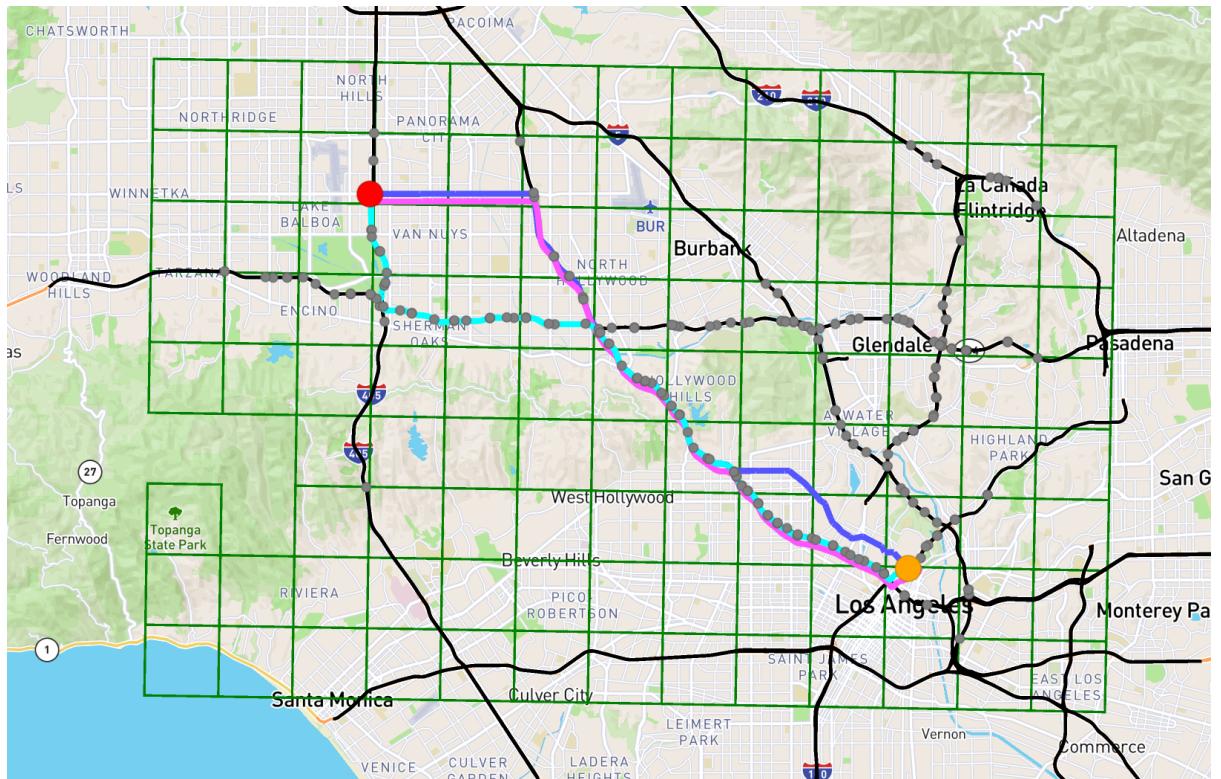
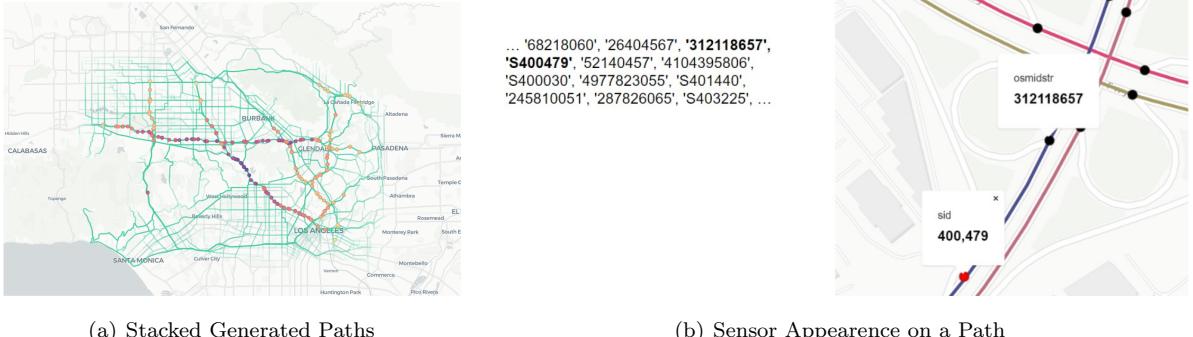


Figure 3.1: The A* algorithm is utilized to generate travel paths between the origin (red) and the destination (orange), sampled from a grid pair, with different costs of using the freeway: the ideal shortest path (blue), and paths that make greater use of the freeway (pink, cyan). The gray markers represent the traffic sensors, which do not necessarily appear on the generated travel paths (in METR-LA).

sensors are installed on one-way freeways where sensors can be reached in consecutive sequences. As a result, the distance matrix is directed, implying that $\text{dist}(v_i, v_j) \neq \text{dist}(v_j, v_i)$.

To construct the adjacency matrix, we apply a Gaussian filter to the distance values. The distance-based proximity matrix between sensors v_i and v_j is computed as $A_{ij}^{(D)} = \exp\left(-\frac{\text{dist}(v_i, v_j)^2}{\sigma^2}\right)$ if $\text{dist}(v_i, v_j) < \kappa$ else 0. While previous works [Li et al., 2018, Yu et al., 2018] leveraged standard deviation ¹ for σ and encountered ambiguity in selecting κ , we consider the specific characteristics of the traffic data. Taking into account the average traffic speed of approximately 60 mph (see Tab. 3.1) and an average distance traveled of 5 miles every 5 minutes (1 time-step), we set σ to 5 miles. Furthermore, we choose κ to be 80 miles, representing the maximum distance that can be covered within one hour (12 time steps), which aligns with the observed maximum speed. As demonstrated, the justification for σ and κ is considerably clearer compared to that of DCRNN.

¹The measured standard deviations of distances are 4.97 miles (METR-LA), 3.93 miles (PEMS-BAY), and 6.92 miles (PEMSD7) respectively. However, it is important to note that the specific σ value can significantly fluctuate depending on the measured distances between sensors, which can potentially challenge the previous approach.



(a) Stacked Generated Paths

(b) Sensor Appearance on a Path

Figure 3.2: (a) Stacked visualization of generated travel paths. (darker color – frequency of the sensor appearance) (b) A travel path contains OSM node IDs and sensor IDs like a sentence.

Figure 3.2 (b) illustrates that multiple sensors can appear along a single generated path. In order to incorporate generated travel trajectories, we calculate the co-occurrence-based [Strehl and Ghosh, 2002] adjacency matrix $A_{ij}^{(S)}$, which measures the likelihood of paths between sensor nodes as follows:

$$A_{ij}^{(S)} = \frac{\# \text{ paths } v_i, v_j \text{ co-appear in } \mathcal{M}^{(Gen)}}{\sqrt{\# \text{ paths } v_i \text{ appears} \times \# \text{ paths } v_j \text{ appears in } \mathcal{M}^{(Gen)}}} \quad (3.2)$$

To obtain the final adjacency matrix to be used for graph convolution, we apply element-wise multiplication of the distance matrix and the co-occurrence matrix as $\mathbf{A} = \mathbf{A}^{(D)} \odot \mathbf{A}^{(S)}$.

3.2.2 Sensor Embedding

Each traffic sensor is situated within a unique built environment, resulting in distinct meanings in the actual traffic speed values. However, obtaining reliable sensor metadata to understand these variations is currently limited. To overcome this challenge, we adopt a similar strategy as described in [Guo et al., 2021], which involves the use of D -dimensional sensor embeddings (SE) generated through one-hot encoding of the N sensors. By incorporating these sensor embeddings, into the input of the encoder and decoder of our models, we can account for the individual characteristics of each sensor.

3.2.3 Activity Embedding

Urban human activity, driven by diverse travel purposes, significantly contributes to traffic congestion [Bowman and Ben-Akiva, 2001, Bhat et al., 2004]. To capture the temporal variations in human activity, we construct the activity frequency based on a weekly pattern derived from the National Household Travel Survey [U.S. Department of Transportation, 2017], as depicted in Fig. 3.3. This allows us to create a representation $H_{t-P+1, \dots, t+Q} \in \mathbb{R}^{K_H}$ that captures the estimated human activities for households at timestamp $t - P + 1, \dots, t + Q$. Subsequently, we first normalize the activity frequency with standard deviation, and is transformed into a D -dimensional activity embedding using a two-stacked dense layer followed by a normalization layer. To incorporate activity embedding into our models, we include $AE_{t-P+1, \dots, t}$ to the input for the encoder, and $AE_{t+1, \dots, t+Q}$ to the input for the decoder by addition, along with sensor embeddings. This allows our models to leverage the contextual activity information in both the encoding and decoding stages.

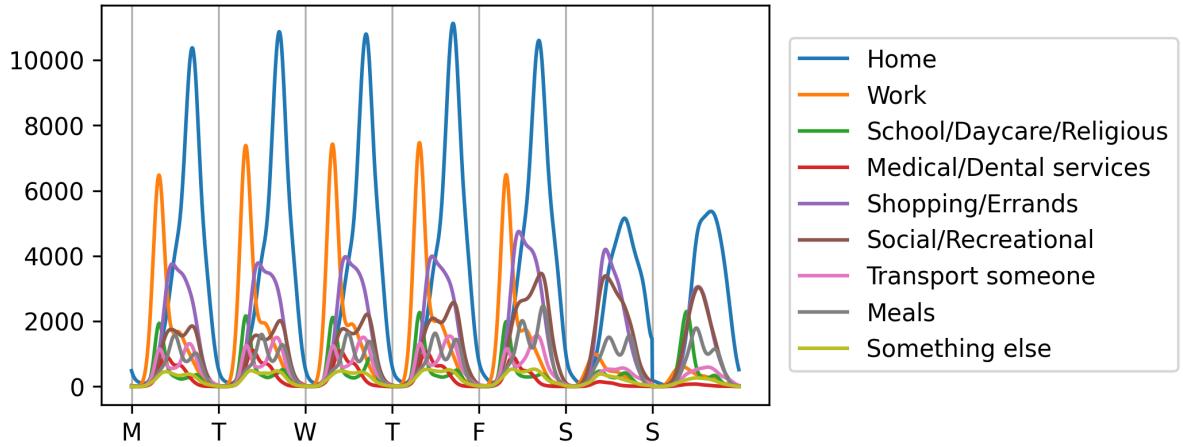


Figure 3.3: Urban human activity frequencies from the National Household Travel Survey for Activity Embedding.

3.2.4 Deep Neural Network

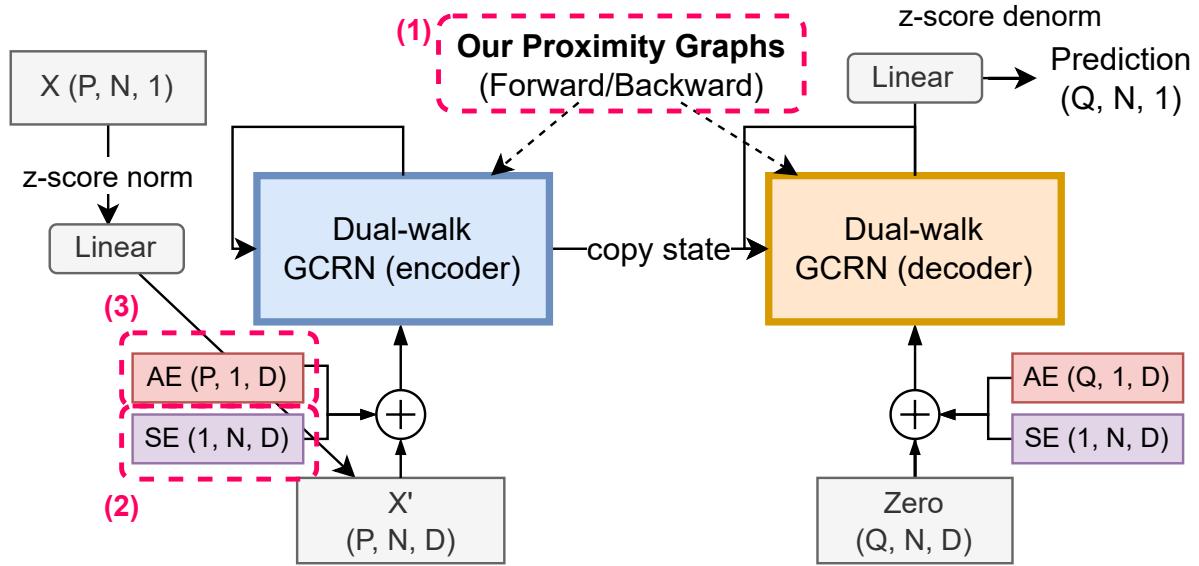


Figure 3.4: Model Architecture (UA-GCRN)

In this section, we present UA-GCRN and UA-GCTransformer as the fundamental models that embody our proposed approach. However, variations of our models, such as UA-LSTM and UA-Transformer, can be implemented without utilizing graph convolution while still considering sensor and activity embedding. To leverage the constructed graph, we introduce dual-walk graph convolution. Additionally, we explore the application of dual-walk graph convolution in two different temporal deep learning methods: recurrent neural network (RNN) and Transformer.

Dual-walk Graph Convolution

We utilize a dual-walk graph convolution approach that combines both diffusion and reverse processes. This involves performing a multi-graph convolution using forward walk, backward walk, and the identity

matrix. The motivation behind employing dual-walk convolution is to address traffic congestion that can occur in both directions due to traffic waves [Daganzo, 1994]. The dual-walk graph convolution is equivalent to a single-step dual-walk diffusion convolution, which was initially proposed in [Li et al., 2018], expressed as follows:

$$g_{\theta_G} Z^{(l+1)} = [\theta_1(\mathbf{D}_{out}^{-1}\mathbf{A}) + \theta_2(\mathbf{D}_{in}^{-1}\mathbf{A}^T) + \theta_0(I)]Z^{(l)} \quad (3.3)$$

Here, $\theta_0, \theta_1, \theta_2$ represent trainable variables, and \mathbf{D}_{out} and \mathbf{D}_{in} are out-degree and in-degree diagonal matrices of \mathbf{A} , respectively. Additionally, we can efficiently perform sparse-matrix computations as our adjacency matrix is sparse.

Graph Convolutional Recurrent Network

We apply dual-walk graph convolutional on GCRN [Seo et al., 2018b, Li et al., 2018] as illustrated in Fig. 3.4. Our UA-GCRN module is equivalent to single-step dual-walk diffusion of DCRNN with sensor and activity embedding. The reason for employing a single-step instead of a multi-step approach, is due to the incorporation of well-engineered sensor connectivity within our graph structure, rendering the multiple diffusion unnecessary. This is experimentally demonstrated in Sec. 3.4.2. Moreover, a single GCRN module still can accumulate graph convolution of each time step to enable multi-step prediction.

Graph Convolutional Transformer

The Transformer architecture [Vaswani, 2017] has achieved remarkable performance in language modeling tasks, leading to various attempts to adapt its structure for other domains. However, recent studies have relied on learnable positional encoding or learnable graph computation techniques [Zheng et al., 2020, Cai et al., 2020, Shao et al., 2022, Jiang et al., 2023]. Interestingly, no existing model has successfully demonstrated the effectiveness of a basic Transformer on traffic prediction, without graph self-learning or modification in positional encodings. In our UA-GCTransformer model, we adopt the original Transformer architecture and utilize sinusoidal positional encoding to distinguish input and output sequence orders as Fig. 3.5. Additionally, we incorporate dual-walk graph convolution in each encoder and decoder layer, similarly in [Guo et al., 2021]. This approach explores the potential of language modeling while leveraging the power of graph convolution.

3.3 Experimental Setting

3.3.1 Data Description and Preprocessing

We provide a description of the datasets used in our study and the corresponding preprocessing steps. Table 3.1 presents the statistics of datasets, including the number of nonzero weights (NNZ) in the adjacency matrix, and the mean betweenness centrality of our adjacency graph.

Traffic Datasets

We utilize three well-known traffic datasets: METR-LA [Li et al., 2018], PEMS-BAY [Li et al., 2018], and PEMSD7 [Yu et al., 2018]. These datasets contain information about traffic speeds recorded by sensors, as well as the original sensor adjacency matrix provided by the authors.

Table 3.1: Data statistics (B.C.: Normalized Betweenness Centrality). *PEMSD7 only contains weekdays.

	METR-LA	PEMS-BAY	PEMSD7*
# sensors (N)	207	325	228
Mean (mph)	54 (± 20)	62 (± 10)	59 (± 13)
Data size	34,249	52,093	12,652
Start time	Mar/1/2012	Jan/1/2017	May/1/2012
End time	Jun/30/2012	May/31/2017	June/30/2012
# OSM roads	75,046	36,987	122,201
$N_H^{(\text{Grid})} \times N_W^{(\text{Grid})}$	9×13 (2mi.)	9×9 (2mi.)	8×12 (3mi.)
$ \mathcal{M}^{(Gen)} $	105,361	46,205	66,510
Legacy Adj. NNZ	1,722 (4.0%)	2,694 (2.6%)	8,100 (15.6%)
Our Adj. NNZ	8,575 (20.%)	12,628 (12.0%)	7,135 (13.7%)
Mean B.C. Ours	3.04×10^{-3}	2.48×10^{-3}	3.32×10^{-3}

Open Street Map (OSM) Dataset

To accurately match the sensor locations to the corresponding roads, we leverage Open Street Map[Ope,] data for the regions covered by the METR-LA, PEMS-BAY, and PEMSD7 datasets. We observed that the locations of some sensors do not align precisely with the OSM roads. In such cases, we updated the latitude, longitude, and freeway details of those sensors using the Caltrans Performance Measurement System (PeMS) [PEM,].

Urban Activity Dataset

To incorporate urban human activity information, we extracted data from the National Household Travel Survey [U.S. Department of Transportation, 2017]. This dataset contains 828,438 travel surveys that include information about travel start and end times, as well as the mode of transportation (including car usage). We constructed an activity frequency histogram with a 5-minute resolution and smoothed pattern with a Gaussian filter ($\sigma=2$) and used it as an input for activity embedding in our model. The number of activity categories is $K_H = 9$ and is described in Fig. 3.3.

Travel Path Generation

We conduct the travel path generation as illustrated in Sec. 3.2.1. We partition the area around the sensors into square grids of 2-3 miles, including padding, resulting in $N_H^{(\text{Grid})} \times N_W^{(\text{Grid})}$ grids. For $(N_H^{(\text{Grid})} \times N_W^{(\text{Grid})})^2$ pairs of grids, we attempt to generate a travel path using the A* algorithm. We perform this process 5 times with 3 different freeway costs (1.0, 0.9, 0.8 multiplied to freeway road length) for each grid pair as described in Fig. 3.1, resulting in a maximum of 15 roads being created². As a result, we can generate $\mathcal{M}^{(Gen)}$ travel paths to construct our co-occurrence and distance-based adjacency matrix. As an example, the adjacency matrix of PEMSD7 is visualized in Fig. 3.7.

²Note that there can be cases where a path is not established.

3.3.2 Evaluation Setup

In our experiments, we evaluate MAE (Mean Absolute Error), RMSE (Root Mean Square Error), and MAPE (Mean Absolute Percentage Error) at 3, 6, and last (12 in METR-LA, PEMS-BAY, 9 in PEMSD7) step of prediction.

We utilized the following parameter settings: a batch size of 32, a hidden embedding dimension³ D of 64, and the Adam optimizer with an initial learning rate of 0.01. We employed a patience of 5 for early stopping and reduced the learning rate to 1/10 after 2 trials. For the Transformer models, we employed 8 attention heads, a key dimension of 8, a total dimension of 64, and stacked 3 layers.

³Dimension of 64 is a frequently employed choice for DCRNN, GTS, and GMAN.

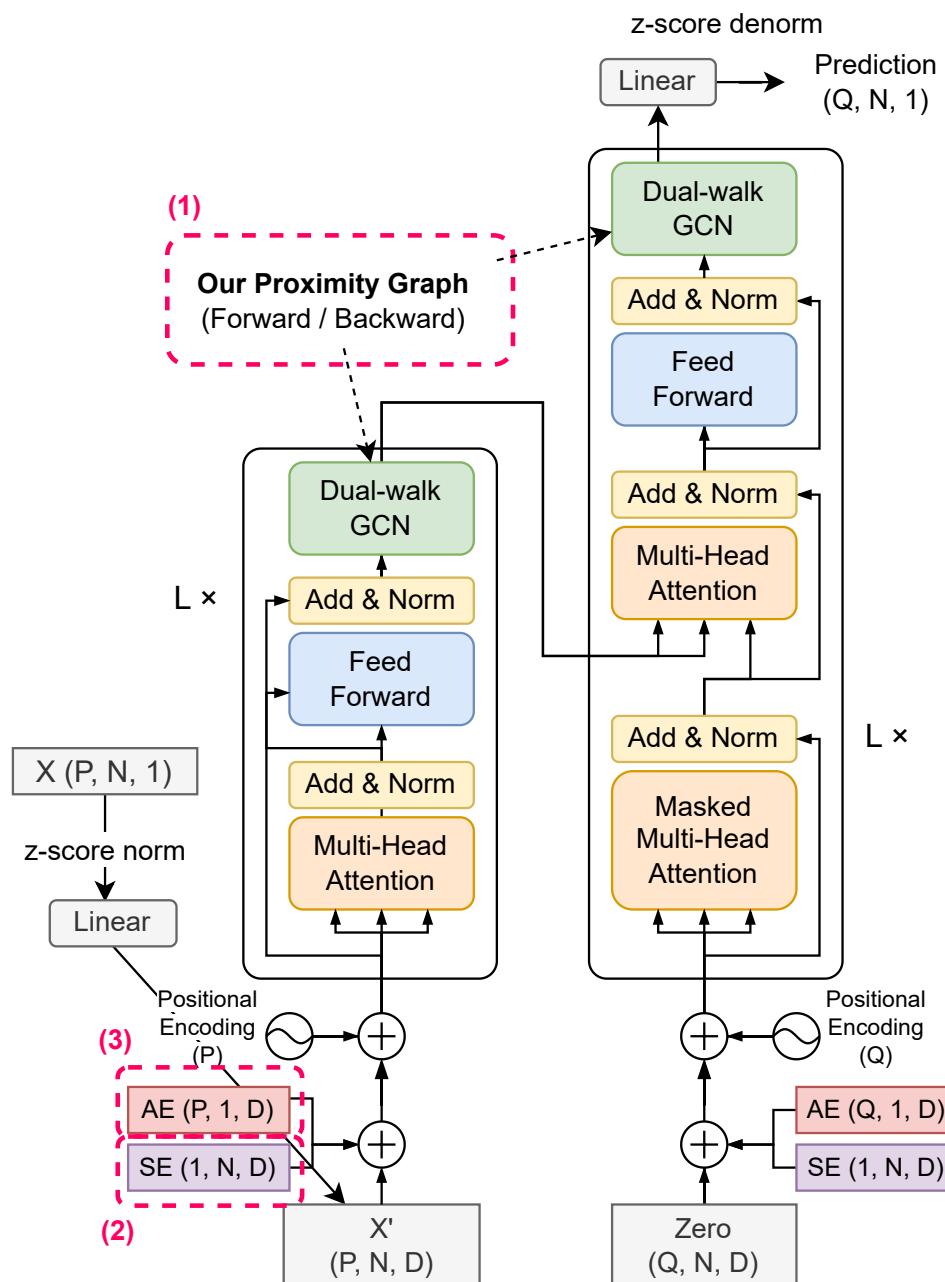


Figure 3.5: Model Architecture (UA-GCTransformer)



Figure 3.6: Traffic sensors (red markers) along with OSM freeways (blue paths) and the corresponding freeways where the traffic sensors are located (green paths) in PEMS-BAY. The partitioned grid is also represented with dark green squares.

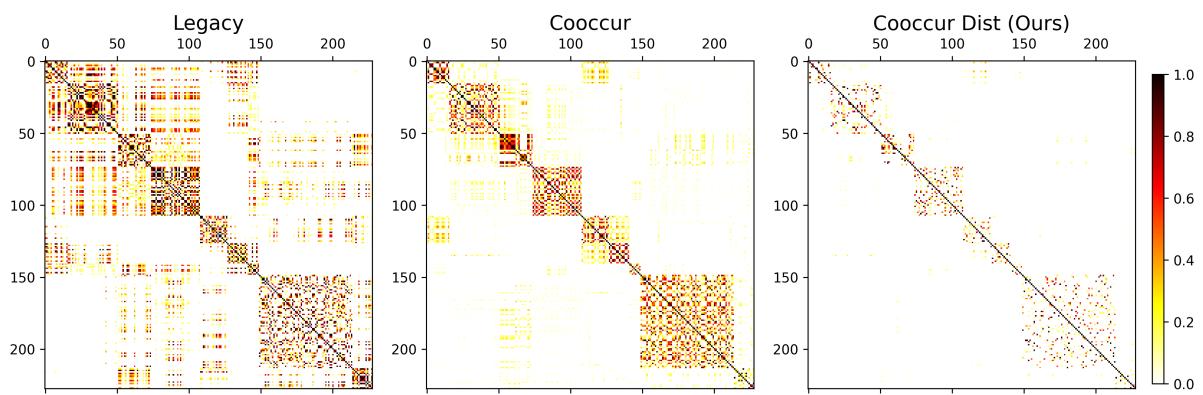


Figure 3.7: Adjacency Matrix Visualization: Legacy, Co-Occurrence Adjacency, and Final Graph in PEMSD7.

Table 3.2: Forecasting error in METR-LA, PEMS-BAY, PEMSD7 datasets. \dagger represents the model leveraging our co-occurrence and distance-based adjacency matrix. * represents the model self-trains the sensor adjacency. Best and second best results are represented as BOLD and underline.

		Metric	Temporal Only				Spatiotemporal								UAGCRNN\dagger	
			LastRepeat	LSTM	TF	UA-LSTM	UA-TF	DCRNN	DCRNN \dagger	GITS*	STGCN	GWNet*	GMAN*	STEP*		
METR-LA	15 min	MAE	4.02	3.09	3.07	2.82	2.81	2.77	2.67	2.67	2.88	2.69	2.77	2.61	2.64	<u>2.63</u>
		RMSE	8.69	6.10	6.09	5.56	5.58	5.38	5.21	5.27	5.74	5.15	5.48	4.98	5.09	<u>5.07</u>
		MAPE	9.4%	8.2%	8.1%	7.5%	7.6%	7.30%	6.89%	7.21%	7.62%	6.90%	7.25%	6.60%	6.77%	<u>6.71%</u>
	30 min	MAE	5.09	3.79	3.76	3.19	3.18	3.15	3.06	3.04	3.47	3.07	3.07	2.96	2.97	<u>2.96</u>
		RMSE	11.13	7.66	7.65	6.59	6.61	6.45	6.30	6.25	7.24	6.22	6.34	5.97	6.08	<u>6.04</u>
		MAPE	12.2%	10.7%	10.6%	9.1%	9.1%	8.80%	8.38%	8.41%	9.57%	8.37%	8.35%	7.96%	8.10%	<u>8.08%</u>
	60 min	MAE	6.80	4.90	4.88	3.56	3.54	3.60	3.56	3.46	4.59	3.53	3.40	3.37	<u>3.35</u>	3.34
		RMSE	14.21	9.68	9.67	7.55	7.52	7.59	7.52	7.31	9.40	7.37	7.21	6.99	7.12	<u>7.02</u>
		MAPE	16.7%	14.9%	14.8%	10.6%	10.7%	10.50%	10.15%	9.98%	12.70%	10.01%	9.72%	9.61%	9.68%	<u>9.65%</u>
PEMS-BAY	15 min	MAE	1.60	1.45	1.45	1.32	1.33	1.38	<u>1.29</u>	1.34	1.36	1.30	1.34	1.26	1.30	1.30
		RMSE	3.43	3.16	3.16	2.82	2.85	2.95	2.72	2.83	2.96	2.74	2.82	<u>2.73</u>	2.73	2.76
		MAPE	3.2%	3.0%	3.0%	2.8%	2.8%	2.90%	<u>2.69%</u>	2.82%	2.90%	2.73%	2.81%	2.59%	2.71%	2.75%
	30 min	MAE	2.18	1.98	1.98	1.63	1.63	1.74	1.62	1.66	1.81	1.63	1.62	1.55	<u>1.61</u>	1.61
		RMSE	4.99	4.61	4.61	3.77	3.78	3.97	3.68	3.78	4.27	3.70	3.72	3.58	<u>3.68</u>	3.70
		MAPE	4.7%	4.5%	4.5%	3.7%	3.7%	3.90%	3.62%	3.77%	4.17%	3.67%	3.63%	3.43%	<u>3.62%</u>	3.64%
	60 min	MAE	3.05	2.72	2.71	1.89	1.88	2.07	1.92	1.95	2.49	1.95	<u>1.86</u>	1.79	1.87	1.86
		RMSE	7.01	6.28	6.27	4.41	4.40	4.74	4.45	4.43	5.69	4.52	<u>4.32</u>	4.20	4.37	4.33
		MAPE	6.8%	6.8%	6.7%	4.5%	4.4%	4.90%	4.52%	4.58%	5.79%	4.63%	<u>4.31%</u>	4.18%	4.39%	4.36%
PEMSD7	15 min	MAE	2.49	2.35	2.37	2.13	2.13	2.21	2.10	2.21	2.25	2.31	2.30	2.09	2.05	<u>2.06</u>
		RMSE	4.65	4.48	4.51	4.03	4.12	4.21	3.98	4.16	4.04	4.44	4.39	3.99	3.87	<u>3.93</u>
		MAPE	5.7%	5.5%	5.5%	5.1%	5.1%	5.14%	4.91%	5.15%	5.26%	5.41%	5.66%	5.00%	4.85%	<u>4.86%</u>
	30 min	MAE	3.51	3.31	3.33	2.71	2.69	3.01	2.75	2.95	3.03	3.26	2.71	2.66	<u>2.61</u>	2.59
		RMSE	6.77	6.49	6.53	5.37	5.49	5.96	5.45	5.74	5.70	6.41	5.35	5.37	5.20	<u>5.22</u>
		MAPE	8.3%	8.1%	8.1%	6.9%	6.9%	7.43%	6.85%	7.43%	7.33%	8.11%	6.87%	6.80%	<u>6.56%</u>	6.50%
	45 min	MAE	4.31	4.05	4.10	3.01	2.98	3.59	3.19	3.47	3.57	4.63	2.99	2.95	<u>2.92</u>	2.90
		RMSE	8.32	7.89	7.99	6.10	6.13	7.14	6.39	6.78	6.77	8.81	5.94	6.03	5.90	<u>5.91</u>
		MAPE	10.4%	10.3%	10.3%	7.9%	7.8%	9.18%	8.24%	9.06%	8.69%	12.40%	7.70%	7.74%	<u>7.58%</u>	7.48%

3.4 Results

3.4.1 Performance Comparison

We have selected several baselines for comparison with our proposed UAGCRN, UAGCTransformer. The baselines include⁴: Last Repeat, LSTM, Transformer[Vaswani, 2017], DCRNN[Li et al., 2018], GTS[Shang et al., 2021], STGCN[Yu et al., 2018], Graph Wavenet (GWNet) [Wu et al., 2019], GMAN[Zheng et al., 2020], STEP[Shao et al., 2022].

Additionally, we also compare our proposed models with two variants: UA-LSTM and UA-Transformer. These variants leverage activity and sensor embeddings but do not incorporate graph convolutions, enabling us to evaluate the impact of graph utilization.

Forecasting error

Tab. 3.2 presents the results of comparing various baseline models with our proposed models (UAGCRN and UAGCTransformer). On the header of the table, † represents the model leveraging our co-occurrence and distance-based adjacency matrix, and * represents the model self-learns the sensor adjacency, which are GTS, GWNet, GMAN, and STEP.

Overall, the proposed models, UAGCRN and UAGCTransformer, consistently outperform the major spatiotemporal baselines across all three datasets and various time intervals. Moreover, these models outperform UALSTM and UATransformer, indicating that incorporating graph convolutions significantly improves the accuracy of traffic forecasting models.

We further evaluate the effectiveness of our graph construction approach by comparing DCRNN, DCRNN†, and GTS⁵, which share the same architecture. Notably, we observed that DCRNN† outperforms GTS, particularly on the PEMS-BAY and PEMSD7 datasets. The superior performance of DCRNN† can be attributed to the fact that GTS considers all potential sensor connections, often resulting in biased predictions. We also noticed this issue of trainable graph adjacency in GWNet, as it exhibits significant errors on the PEMSD7 dataset. The sensor networks in the PEMSD7 dataset have higher betweenness centrality (Tab. 3.1), indicating that the graph structure provides more valuable information compared to other datasets. These complexities in sensor adjacency may also contribute to the lower performance of the STEP model compared to our approach, as STEP relies on a data-driven approach that may not accurately capture the intricate sensor relationships.

Moreover, our proposed UA approach, which includes SE and AE components, significantly improves the performance of purely temporal models (LSTM and TF). UA-LSTM and UA-Transformer surpass other spatiotemporal baselines such as DCRNN, GTS, STGCN, GWNet, and GMAN on the PEMS-BAY and PEMSD7 datasets. This observation suggests that by incorporating $(P + Q) \times N$ types of inputs, which include both sensor index and urban activity context, our models are able to distinguish between different sensor inputs and capture distinct activity contexts. This comprehensive input representation empowers our models to generate accurate predictions for multi-step traffic forecasting tasks.

However, we observe limited improvement of Transformer over LSTM and UAGCTransformer over UAGCRN. This can be attributed to the fact that traffic prediction involves relatively short time-series

⁴ Although we considered Traffic Transfomer[Cai et al., 2020] for its state-of-the-art performance, we encountered difficulties in finding a reliable dataset or test logs. We assume there might be confusion in metrics such as averaging over multiple time steps.

⁵Results from [Shao et al., 2022] due to issues with GTS: <https://github.com/chaoshangcs/GTS/issues>

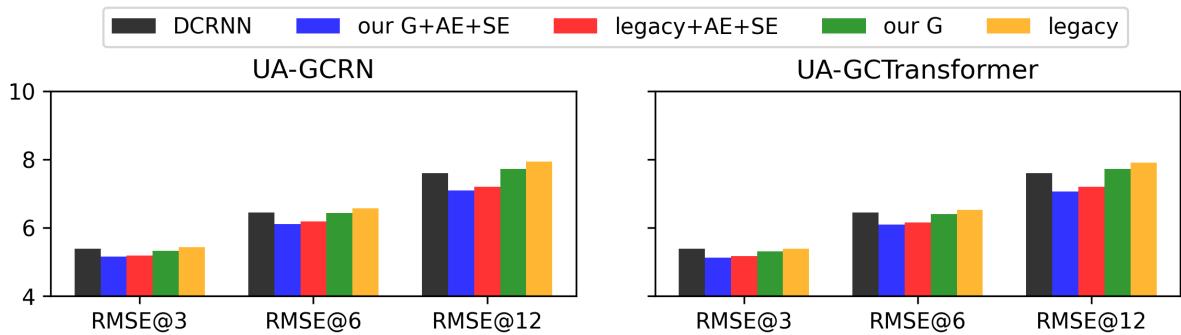


Figure 3.8: METR-LA (UA-GCRN, UA-GCTransformer)

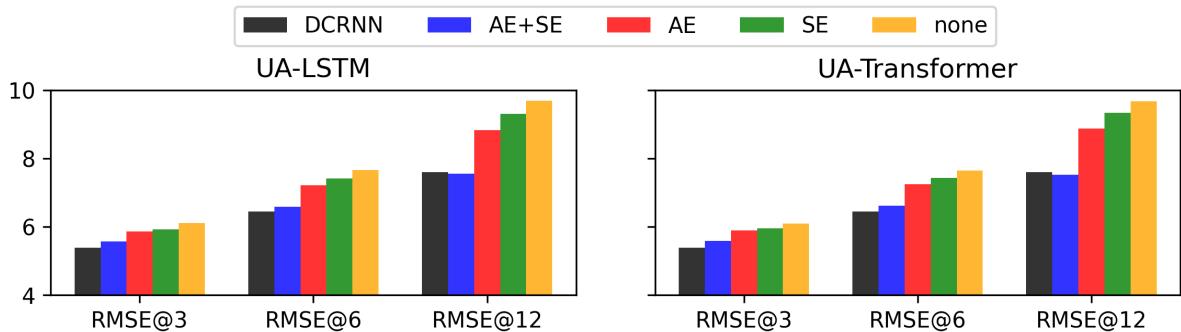


Figure 3.9: METR-LA (UA-LSTM, UA-Transformer)

Figure 3.10: Ablation Test (RMSE) of our modules – Our Graph(G), SE, A on METR-LA.

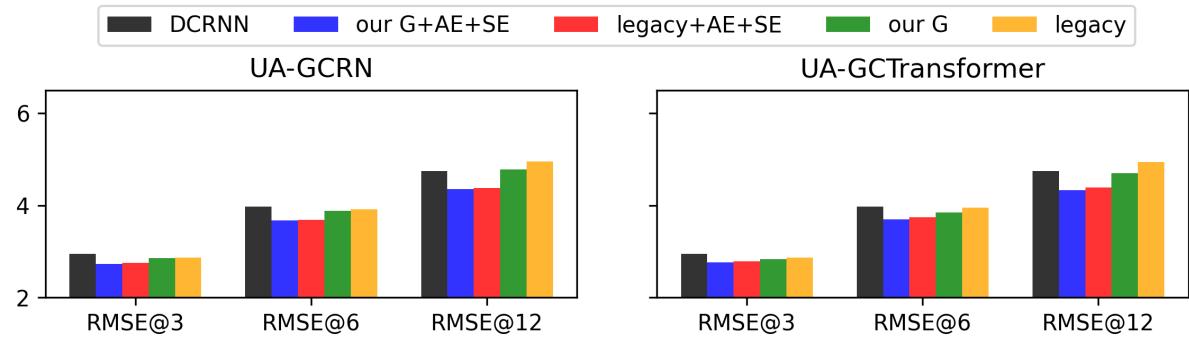


Figure 3.11: PEMs-BAY (UA-GCRN, UA-GCTransformer)

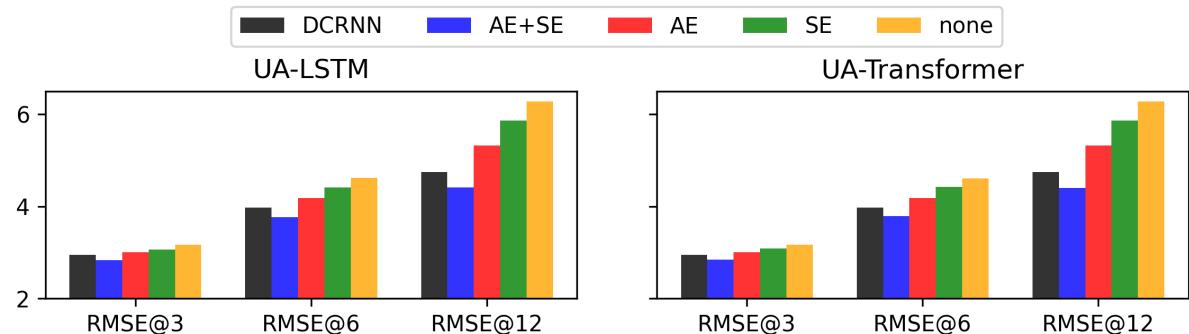


Figure 3.12: PEMs-BAY (UA-LSTM, UA-Transformer)

Figure 3.13: Ablation Test (RMSE) of our modules – Our Graph(G), SE, A on PEMs-BAY.

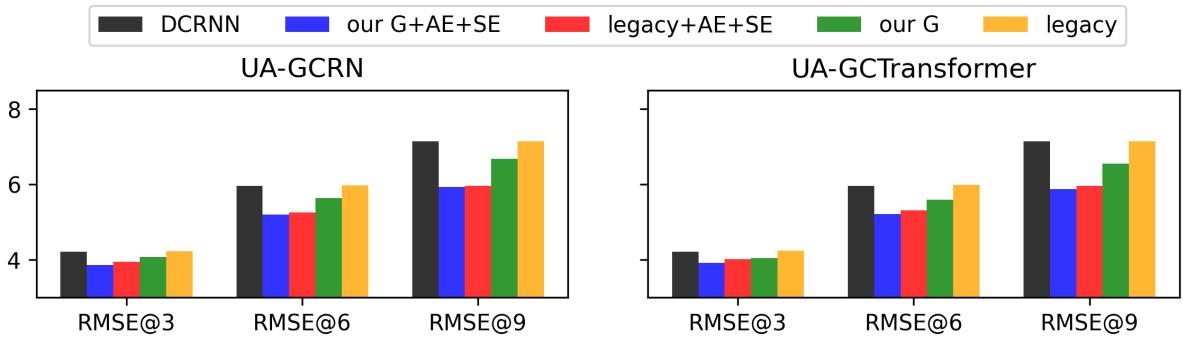


Figure 3.14: PEMSD7 (UA-GCRN, UA-GCTransformer)

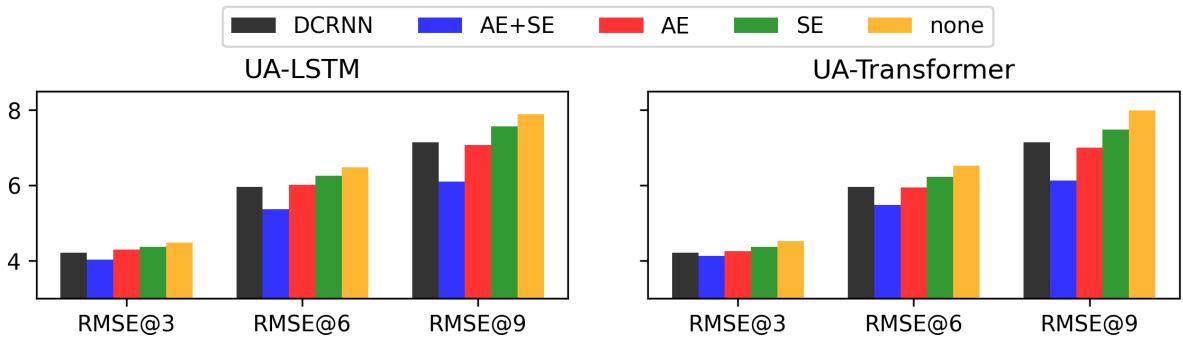


Figure 3.15: PEMSD7 (UA-LSTM, UA-Transformer)

Figure 3.16: Ablation Test (RMSE) of our modules – Our Graph(G), SE, A on PEMSD7.

steps, unlike in the case of large language models, where Transformers excel. Consequently, RNN models continue to perform well in this domain.

Although our proposed models, UAGCRN \dagger and UAGCTF \dagger , are outperformed by STEP on the METR-LA and PEMS-BAY datasets, STEP utilizes very long patches of input (e.g. $P = 228 \times 7$) and complex transformer architecture which allows it to capture more complex and intricate patterns. This approach results in a heavier model that can handle larger contextual information, as it shows better performance in longer timesteps. Despite the performance difference, our models still demonstrate potential for improvement, which will be explained in Sec. 3.4.2. On the other hand, we believe that studying STEP’s architecture and mechanisms can provide valuable insights to advance the state-of-the-art in related models.

Computational Cost

We conducted a comparison of the computational cost for each model in their default settings in Tab. 3.3⁶. In order to ensure a fair comparison we leverage the default settings of each model such as DCRNN with 3 diffusion steps, GMAN with $L = 5$. We were unable to precisely measure the computational cost of STEP[Shao et al., 2022] under the same environment. However, during our experiments of STEP with the PEMSD7 dataset (2.7 times smaller than METR-LA), each epoch took approximately 3.5 minutes to train using 3 TITAN RTX GPUs. The total training time was approximately 5 hours. The result shows that UAGCRN outperforms other models in terms of computational cost and training time.

Table 3.3: Computational cost of METR-LA under the same environment. The number of stacks is $L = 5$ in GMAN and $L = 3$ in UAGCTF \dagger , while DCRNN, UAGCRN \dagger do not have stacked architecture ($L = 1$).

	DCRNN	GMAN	UAGCRN \dagger	UAGCTF \dagger
# Params	353,025	714,049	174,401	842,177
Train (m:s/ep.)	2:35	4:39	42s	4:26
Total Epochs	26	19	24	17
Total train time	1:12:03	1:34:41	0:18:36	1:21:04

3.4.2 Ablation Study

Effectness of Graph, AE, SE

The results of the ablation test for each module are presented in Fig. 3.16. Specifically, Fig. 3.8,3.11,3.14 demonstrate the performance improvement of UA-GCRN and UA-GCTransformer when using our graph compared to the legacy graph. These results indicate that our graph contains more traffic-related knowledge regarding sensor correlation. Although the enhancement looks marginal when both the SE and AE modules are given under the same conditions, it still highlights the potential for performance improvement in less common situations.

Furthermore, Fig. 3.9,3.11,3.15 showcase the effectiveness of each SE and AE module on temporal-only models, specifically UA-LSTM and UA-Transformer. The performance improves as these modules

⁶Tab. 3.2 is based solely on the original author’s implementation, while Tab. 3.3 is intended for evaluating computational time under same learning framework (TensorFlow2) and GPU (RTX3090), batch size, and early stopping condition.

are integrated. Notably, the impact of the AE module is more significant than the SE module, likely due to traffic patterns exhibiting a stronger correlation with human activity. Additionally, incorporating the SE module to account for sensor spatial heterogeneity further enhances performance. When both SE and AE modules are integrated, the resulting performance surpasses that of DCRNN in the PEMS-BAY and PEMSD7 datasets.

The number of diffusion steps of DCRNN with our graph

Figure 3.17 depicts the results of UADCGRU \dagger obtained by applying the SE and AE to the DCGRU model using our graph while modifying the diffusion steps. In contrast to the original findings discussed in [Li et al., 2018] which suggested a demand for approximately 3 diffusion steps, our results show that increasing the graph connectivity information, along with activity and sensor data, can lead to worse performance. We analyze that vehicles do not follow a random walk pattern, and the vehicle travel pattern is already adequately captured in our constructed graph.

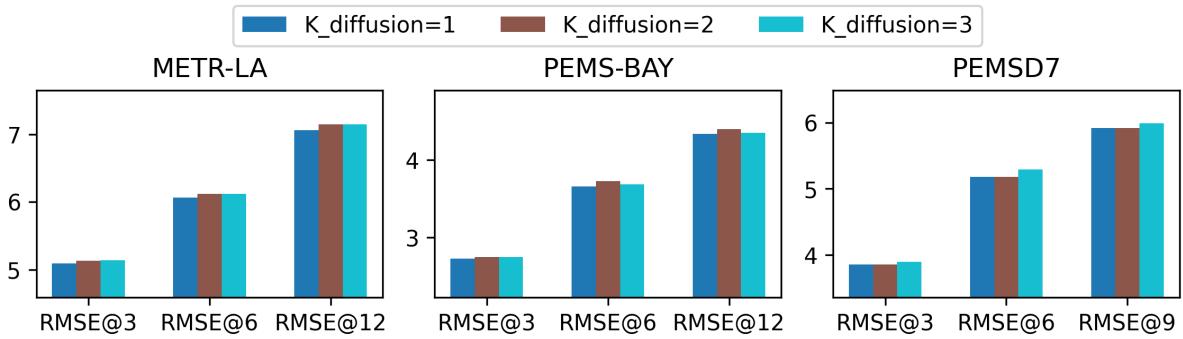


Figure 3.17: Performance degradation in UADCGRU \dagger as the number of diffusion steps (K) increases.

Comparison of Timestamp Embedding and Activity Embedding

Various models have employed different approaches to incorporate timestamp information. For example, in DCRNN, the time of day is included as an additional input channel⁷. In this ablation study, we compared timestamp embedding (TE), which are generated from a vector space $\{0,1\}^{7+12\times 24}$ (one-hot concatenation of weekday and time-of-day) and ingested in a 2-stacked dense layer with a normalization layer to capture weekly and daily periodicity, similar to [Zheng et al., 2020, Jiang et al., 2023], with AE.

Tab. 3.4 shows the comparison results of UAGCRN \dagger with TE or AE, indicating that TE exhibited a slight improvement over AE, performing almost as well as STEP in the METR-LA and PEMS-BAY datasets, and outperforming AE in the PEMSD7 dataset. The performance improvement of the TE over the AE can be attributed to the lack of analysis of localized activity patterns of each city when we estimate human activity frequency which is derived from national surveys. This aspect suggests that future studies should consider accurately inputting AE, such as localized activity estimation considering demographics and urban function.

On the other hand, relying on one-hot timestamp information results in less explainability due to its discrete nature, unlike continuous activity information. Additionally, it may pose limitations in scalability when accounting for seasonal effects in long-term datasets, while our datasets are deal with

⁷Not mentioned in the paper, but in the code: <https://github.com/liyaguang/DCRNN>

Table 3.4: Ablation study of UAGCRN \dagger and UAGCTF \dagger by replacing AE with timestamp embedding (TE). Best and second best results are represented as **BOLD** and underline.

	STEP	UAGCRN \dagger		UAGCTF \dagger	
		TE+SE	AE+SE	TE+SE	AE+SE
METR-LA	MAE3	2.61	<u>2.62</u>	2.64	2.63
	RMSE3	4.98	<u>5.00</u>	5.09	5.07
	MAE6	<u>2.96</u>	2.94	2.97	2.95
	RMSE6	<u>5.97</u>	5.97	6.08	6.04
	MAE12	3.37	3.31	3.35	3.35
	RMSE12	6.99	<u>7.02</u>	7.12	7.02
PEMS-BAY	MAE3	1.26	<u>1.28</u>	1.30	1.28
	RMSE3	<u>2.73</u>	2.69	2.73	2.72
	MAE6	1.55	1.60	1.61	<u>1.59</u>
	RMSE6	3.58	<u>3.63</u>	3.68	3.66
	MAE12	1.79	1.88	1.87	1.86
	RMSE12	4.20	4.38	4.37	<u>4.33</u>
PEMSD7	MAE3	2.09	2.02	2.05	<u>2.04</u>
	RMSE3	3.99	3.81	<u>3.87</u>	3.88
	MAE6	2.66	2.56	2.61	<u>2.57</u>
	RMSE6	5.37	5.14	5.20	<u>5.16</u>
	MAE9	2.95	2.88	2.92	<u>2.89</u>
	RMSE9	6.03	5.88	5.90	<u>5.88</u>

only a few months (Tab. 3.1). Nevertheless, we can still take advantage of the AE-based UAGCRN model for its superior explainability.

3.4.3 Case Study

Fig.3.18 and Fig.3.19 present a case study illustrating the superior performance of UA-GCRN \dagger with our graph, sensor, and activity embeddings. In both cases, the legacy graph includes incorrect connections that cannot be reached from the target sensor, which causes wrong predictions.

In the METR-LA dataset (Fig. 3.18), UA-GCRN \dagger achieves better congestion prediction even without sensor and activity embeddings by accurately establishing connections between roads. This highlights the effectiveness of our approach in constructing the graph, which significantly improves the model’s performance.

Furthermore, in the PEMS-BAY dataset (Fig. 3.19), we observe that UA-GCRN \dagger performs even better when provided with activity input. In this case, a high frequency of work activity is included in the historical sequence, and possible shopping activity in the future sequence, which helps the model there can be consequent congestion in the prediction steps. The results demonstrate the additional benefit of incorporating activity information into the model, further enhancing its performance.

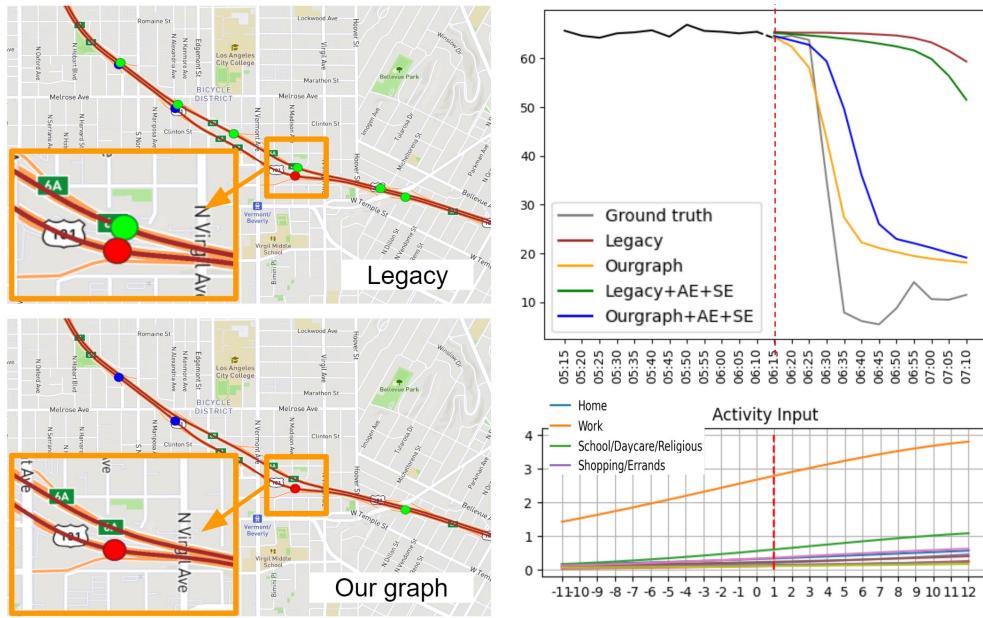


Figure 3.18: METR-LA, outperforms with our graph

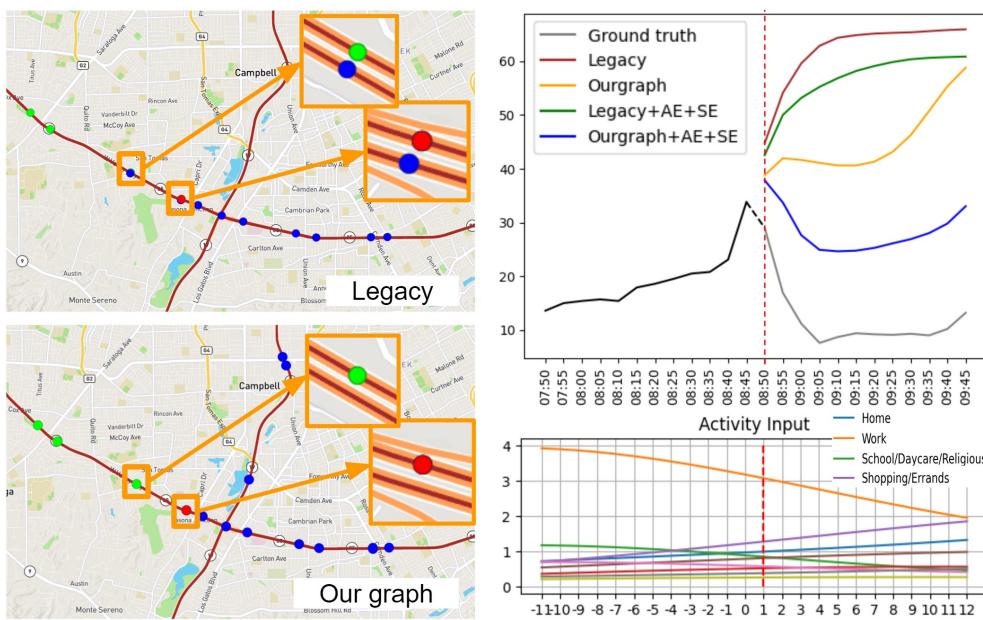


Figure 3.19: PEMS-BAY, outperforms with our graph

3.4.4 Sensor Reactions Based on Activity Input

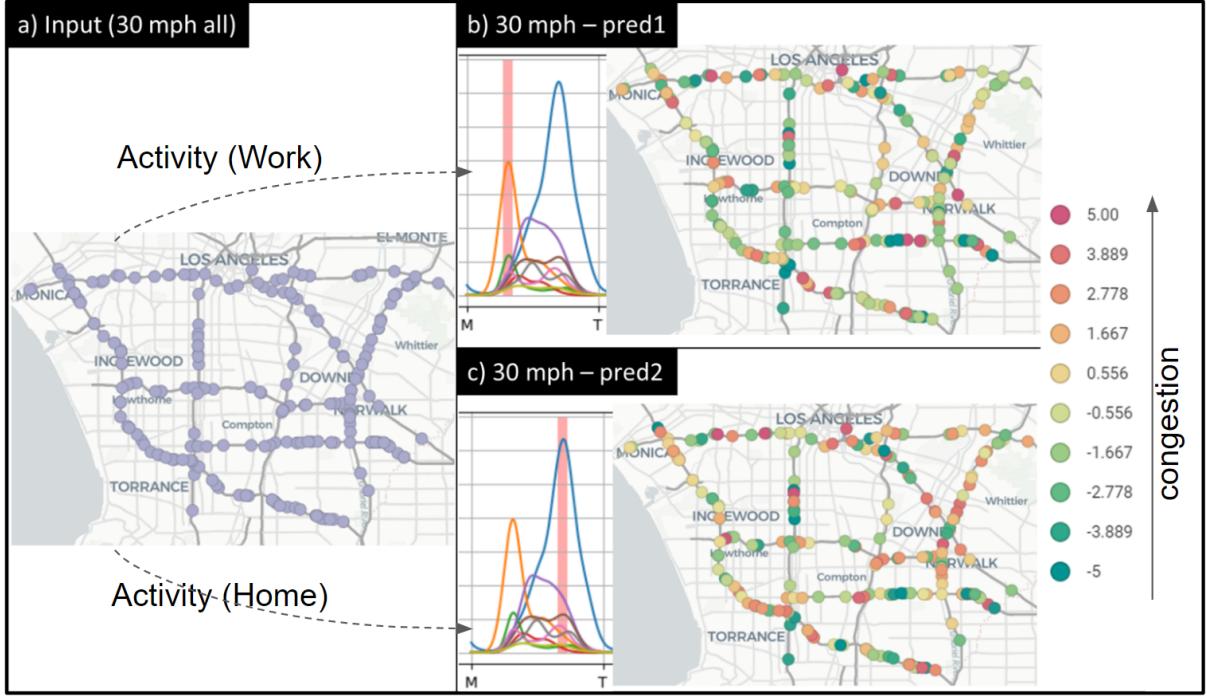


Figure 3.20: Sensor Reactions Based on Activity Information with UAGCRN (Red/Green: more/less congestion)

We examine how sensors react differently when provided with different activity information. We conducted tests by setting all sensors to a speed value of 30 mph during the P sequence while varying the activity input, as illustrated in Fig. 3.20. The choice of 30 mph is for testing whether congestion would increase or alleviate when the road capacity is full. We conducted two predictions of the next 15 min, pred1 and pred2, by providing activity information for the morning rush hour (6:35 to 8:20) and the evening commuting time (16:45 to 18:30), respectively.

Our findings revealed that sensors exhibited different behaviors based on the given activities. This discrepancy is due to varying levels of road utilization associated with specific activities. Notably, even when the same traffic values are given to the model, our model predicted distinct patterns as it had learned the sensor's typical response patterns corresponding to future activities. Overall, these analyses highlight the importance of incorporating sensor embedding while inserting activity information into traffic prediction models, as it leads to a better understanding of sensor reactions and enhances the accuracy of congestion predictions influenced by urban human activity.

Chapter 4. Discussion

4.1 Discussion on Traffic Forecasting

4.1.1 About UAGCRN

Our current research focuses on enhancing activity-based traffic prediction models that consider spatial and temporal factors while incorporating travel purposes. However, there are notable areas for refinement. To enhance the realism of travel paths, it is essential to account for travel demands over time by integrating real-time data such as transportation and social media, enabling accurate inference of road-specific travel demands. Additionally, insights from building use, points of interest, and demographics can deepen our understanding of travel purposes and traffic patterns. Moreover, employing traffic simulation for synthetic data generation offers the potential to uncover nuanced traffic behaviors beyond our current A* algorithm-based approach. Exploring this dynamic route choice behavior using established transportation research and AI techniques could yield more effective predictions. By refining our methodology in these areas, future research can enhance the accuracy and applicability of our models, allowing us to understand better and predict traffic patterns in urban areas.

4.1.2 Challenges in Applying UAGCRN to South Korea

I have dedicated significant efforts to applying various traffic forecasting models, including the seminal DCRNN and GMAN, to Korean data and real-time floating population data in pursuit of practical applications[Han et al., 2024]¹. The proposal of UAGCRN[Han et al., 2023] emerged as a result of questioning the reliability of existing models and re-evaluating their underlying assumptions during a three-year journey of attempting to adapt American models to Korean contexts without critical review.

Through this process, I began to doubt the original DCRNN model, revisited its dataset for validation, and reconsidered the architectural simplicity proposed by alternative models. These reflections culminated in the development of UAGCRN. However, I acknowledge and discuss the limitations that make UAGCRN challenging to implement effectively in South Korea. These include differences in urban dynamics, data characteristics, and infrastructure between South Korea and the U.S., which highlight the need for models specifically tailored to the unique features of Korean urban environments.

1. **High Population and Building Density:** South Korea's limited land area and the high density of buildings and population in urban areas result in exceptionally complex datasets. Unlike in the U.S., models must simulate significantly more movement paths within much smaller, highly congested spaces.
2. **Complex Traffic and Signal Systems:** South Korea has an extensive traffic signal system, a high volume of pedestrians, and a well-developed public transportation network. Features such as Bus Rapid Transit (BRT) lanes add further constraints to the model. In contrast, the U.S. relies on wide lanes and freeway-centric systems, which align better with models based on simpler road structures.

¹An Arxiv Paper initially submitted and rejected from AAAI in 2021.

3. **Diverse and Rapidly Changing Urban Environments:** While planned districts like Gangnam have relatively organized road structures, older areas such as northern Seoul feature irregular streets and numerous alleys. Factors such as narrow alleys, illegal parking, and underground parking lots contribute to sparse cases in simulations, increasing data complexity. Additionally, South Korea's urban infrastructure changes rapidly, exemplified by the introduction of transit systems like the Seoul Station BRT within just a few years. Models must adapt to such dynamic changes.
4. **High-rise Apartments and Commuting Patterns:** The prevalence of high-rise apartments in South Korea necessitates simulating commuting patterns, such as people traveling from residential complexes to office buildings during rush hours. These patterns require finer-grained analysis that differs significantly from those in American cities.
5. **Dynamic Commercial Areas and Tourist Influx:** South Korea experiences rapid changes in commercial districts, influenced by real estate trends, permits for commercial facilities, and fluctuating numbers of tourists. Incorporating these dynamic economic changes into the model requires integrating additional datasets.
6. **Diverse Taxi and Transportation Services:** South Korea has a high number of taxis and a diverse mix of transportation services, introducing additional variables for the model to consider. Especially, citizens in South Korea primarily commute with subways rather than personal vehicles. This complexity necessitates more sophisticated simulations compared to existing models.
7. **Limitations in Pedestrian Simulations:** South Korea's streets are crowded with pedestrians, narrow sidewalks, and frequent instances of jaywalking. Accurately simulating scenarios with concentrated pedestrian activity is essential. Additionally, in alley environments, the shortest path and the preferred path may differ. Failing to account for such discrepancies reduces the model's realism and reliability.
8. **Lack of Real-time Data:** In cities like New York, near-real-time public transportation data is available at intervals as short as five minutes. In contrast, South Korea often relies on statistical data published at hourly intervals, limiting the use of real-time data. This lack of real-time data reflecting human activity poses challenges to improving traffic forecasting performance.
9. **U.S.-centric Model Design:** UAGCRN, like many traffic forecasting models, is primarily designed for the relatively simple road structures and traffic patterns of U.S. freeways. This design creates limitations when applied to South Korea's complex, rapidly evolving urban environments and traffic systems.

To effectively apply UAGCRN to South Korea, the model must be extended to reflect the country's unique urban structures and traffic characteristics. Addressing challenges such as high-density environments, pedestrian-centered mobility patterns, and the scarcity of real-time data requires a novel approach that integrates urban engineering data with real-time human activity data.

4.2 Thoughts on True Future Prediction

“An unchanging world, sufficient as it is, cannot truly be called living. It is merely a world of memories—a completed, closed world. I will refuse it.” – from a movie quote.

The future, I believe, is not something to predict but something to create. Current traffic prediction models typically work with a short-term perspective of around 20 weeks, splitting the data sequentially into 70% for training, 10% for validation, and 20% for testing. This means 14 weeks are used for training, about 2 weeks for validation, and the remaining 4 weeks for testing. Since cities do not often undergo rapid changes, these models can be updated with the latest data each time a prediction is required, thereby achieving significantly better performance than traditional methods like ARIMA, SVM, or RFR.

However, to predict the kinds of transformative changes introduced by entirely new urban developments, as discussed in the introduction, one must constantly grasp people’s desires and engage in conversations to understand them. Most importantly, one must track the flow of money. This could extend to seemingly minor details, such as which lectures by speakers are gaining popularity, what literary works are trending, which music is resonating with audiences, which products or foods are favored, and even which Netflix programs are being widely watched.

Two decades ago, physically labor-intensive workers might have relieved their stress by eating spicy and salty food. However, the younger generation tends to approach food as an art form, valuing not only taste but also presentation and the ambiance of the restaurant. Understanding such phenomena may even require delving into the frustrations of young people dissatisfied with traditional jobs, including the perspectives of unemployed youth, and comprehending the various social issues mentioned earlier in the preface. With this understanding, older generations bear the responsibility of designing new futures for education and industry.

That said, the youth also have responsibilities. If the existing world does not satisfy them, they must create their own. Traditional education systems often evaluated students based on how well they absorbed knowledge, using numerical scores as the standard for their value. As a result, someone might be praised and put on the path to success simply for scoring better on a test, while others might face despair and prematurely label their lives as failures. But isn’t life too precious to live day by day in an unchanging routine, resigning oneself to the belief that they were born only to endure this? No matter how successful a person may seem, they too have fears. Moreover, to have nothing is to wield an incredible weapon—the lack of anything to lose.

Two hundred years ago, pursuing a dream might have meant risking one’s life. But thanks to the sacrifices of unknown ancestors, we now live in a nation where dreams can be pursued without fear of losing one’s life. Furthermore, rapid economic growth has created a society where people no longer starve and benefit from incredibly convenient services, infrastructure, and welfare. How fortunate is that!

When you are determined, you find ways; when you are not, you find excuses. Money, too, is merely an illusion of numbers—a means to achieve dreams. I hope the younger generation will appreciate the country that their predecessors built, refrain from envy or jealousy, and recognize that all humans are inherently equal. Rather than conforming to others’ expectations, trends, or visible societal standards, I urge them to discover who they truly are, what they can do in their own positions, and what brings them happiness. May they never lose confidence or hope and live for their own happiness.

Chapter 5. Conclusion

This dissertation underscores the significance of incorporating real-world knowledge of urban human activity into spatiotemporal traffic prediction models. By addressing key challenges such as accurate graph construction, sensor heterogeneity, and human activity-based inference, this work presents a novel, integrated framework that enhances the predictive power of traffic models.

The proposed approach introduces innovative components, including realistic travel path generation using the A* algorithm, co-occurrence and distance-based sensor connectivity measures, sensor-specific one-hot encodings, and embeddings that capture human activity dynamics. These components are seamlessly integrated into graph-convolution-based spatiotemporal deep learning architectures, ensuring a more nuanced and accurate representation of traffic patterns.

Through comprehensive experiments on real-world datasets, the proposed method demonstrates superior performance over existing baselines, achieving state-of-the-art results. These findings reveal the substantial impact of human activity insights on traffic prediction accuracy and contribute to a deeper understanding of the interplay between urban vibrancy and traffic flow.

By bridging the gap between human mobility behaviors and data-driven traffic modeling, this research lays the groundwork for future advancements in spatiotemporal forecasting. Potential extensions include the integration of real-time urban data, further refinements of graph-based methods, and exploration of adaptive models that can accommodate evolving urban environments.

Ultimately, this dissertation not only advances the technical landscape of traffic prediction but also offers practical insights for urban planning, intelligent transportation systems, and smart city initiatives, fostering more efficient and sustainable urban mobility solutions.

Bibliography

- [PEM,] California Performance Measurement System (PeMS). <https://pems.dot.ca.gov/>. Accessed: June 3, 2023.
- [Ope,] Openstreetmap. <https://www.openstreetmap.org>. Accessed: June 3, 2023.
- [Bhat et al., 2004] Bhat, C. R., Guo, J. Y., Srinivasan, S., and Sivakumar, A. (2004). Comprehensive econometric microsimulator for daily activity-travel patterns. *Transportation Research Record*, 1894(1):57–66.
- [Bischof et al., 2022] Bischof, R., Hansen, N. R., Nyheim, Ø. S., Kisen, A., Prestmoen, L., and Hau-gaasen, T. (2022). Mapping the “catscape” formed by a population of pet cats with outdoor access. *Scientific reports*, 12(1):5964.
- [Bowman and Ben-Akiva, 2001] Bowman, J. L. and Ben-Akiva, M. E. (2001). Activity-based disaggregate travel demand model system with activity schedules. *Transportation research part a: policy and practice*, 35(1):1–28.
- [Cai et al., 2020] Cai, L., Janowicz, K., Mai, G., Yan, B., and Zhu, R. (2020). Traffic transformer: Capturing the continuity and periodicity of time series for traffic forecasting. *Transactions in GIS*, 24(3):736–755.
- [Chung et al., 2014] 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*.
- [contributors, nd] contributors, W. (n.d.). Simcity. Accessed: 2024-11-19.
- [Daganzo, 1994] Daganzo, C. F. (1994). The cell transmission model: A dynamic representation of highway traffic consistent with the hydrodynamic theory. *Transportation research part B: methodological*, 28(4):269–287.
- [Entertainment, nd] Entertainment, I. (n.d.). Simcity: Traffic. Accessed: 2024-11-19.
- [Fang et al., 2021] Fang, Z., Long, Q., Song, G., and Xie, K. (2021). Spatial-temporal graph ode networks for traffic flow forecasting. In *Proceedings of the 27th ACM SIGKDD conference on knowledge discovery and data mining*, pages 364–373.
- [Forrester, 1969] Forrester, J. W. (1969). *Urban Dynamics*. Pegasus Communications, Waltham, MA.
- [Forrester, 1971] Forrester, J. W. (1971). *World Dynamics*. Wright-Allen Press, Cambridge, MA.
- [Guo et al., 2019] Guo, S., Lin, Y., Feng, N., Song, C., and Wan, H. (2019). Attention based spatial-temporal graph convolutional networks for traffic flow forecasting. In *Proceedings of the AAAI conference on artificial intelligence*, volume 33, pages 922–929.
- [Guo et al., 2021] Guo, S., Lin, Y., Wan, H., Li, X., and Cong, G. (2021). Learning dynamics and heterogeneity of spatial-temporal graph data for traffic forecasting. *IEEE Transactions on Knowledge and Data Engineering*, 34(11):5415–5428.

- [Han et al., 2024] Han, S., An, J., and Lee, D. (2024). Spatio-temporal road traffic prediction using real-time regional knowledge. *arXiv preprint arXiv:2408.12882*.
- [Han et al., 2023] Han, S., Park, Y., Lee, M., An, J., and Lee, D. (2023). Enhancing spatio-temporal traffic prediction through urban human activity analysis. In *Proceedings of the 32nd ACM International Conference on Information and Knowledge Management*, pages 689–698.
- [Hart et al., 1968] Hart, P. E., Nilsson, N. J., and Raphael, B. (1968). A formal basis for the heuristic determination of minimum cost paths. *IEEE transactions on Systems Science and Cybernetics*, 4(2):100–107.
- [Jang et al., 2016] Jang, E., Gu, S., and Poole, B. (2016). Categorical reparameterization with gumbel-softmax. *arXiv preprint arXiv:1611.01144*.
- [Jiang et al., 2023] Jiang, J., Han, C., Zhao, W. X., and Wang, J. (2023). Pdformer: Propagation delay-aware dynamic long-range transformer for traffic flow prediction. In *Proceedings of the AAAI conference on artificial intelligence*, volume 37, pages 4365–4373.
- [Li et al., 2018] Li, Y., Yu, R., Shahabi, C., and Liu, Y. (2018). Diffusion convolutional recurrent neural network: Data-driven traffic forecasting. In *6th International Conference on Learning Representations, ICLR 2018, Vancouver, BC, Canada, April 30 - May 3, 2018, Conference Track Proceedings*.
- [Liang et al., 2018] Liang, Y., Ke, S., Zhang, J., Yi, X., and Zheng, Y. (2018). Geoman: Multi-level attention networks for geo-sensory time series prediction. In *IJCAI*, volume 2018, pages 3428–3434.
- [Maddison et al., 2016] Maddison, C. J., Mnih, A., and Teh, Y. W. (2016). The concrete distribution: A continuous relaxation of discrete random variables. *arXiv preprint arXiv:1611.00712*.
- [Oord, 2016] Oord, A. v. d. (2016). Wavenet: A generative model for raw audio. *arXiv preprint arXiv:1609.03499*.
- [Schopenhauer, 1833] Schopenhauer, A. (1833). *Die Welt als Wille und Vorstellung*. F.A. Brockhaus, Leipzig, Germany, 2 edition.
- [Seo et al., 2018a] Seo, Y., Defferrard, M., Vandergheynst, P., and Bresson, X. (2018a). Structured sequence modeling with graph convolutional recurrent networks. In *Neural Information Processing: 25th International Conference, ICONIP 2018, Siem Reap, Cambodia, December 13-16, 2018, Proceedings, Part I 25*, pages 362–373. Springer.
- [Seo et al., 2018b] Seo, Y., Defferrard, M., Vandergheynst, P., and Bresson, X. (2018b). Structured sequence modeling with graph convolutional recurrent networks. In *Neural Information Processing: 25th International Conference, ICONIP 2018, Siem Reap, Cambodia, December 13-16, 2018, Proceedings, Part I 25*, pages 362–373. Springer.
- [Shang et al., 2021] Shang, C., Chen, J., and Bi, J. (2021). Discrete graph structure learning for forecasting multiple time series. In *9th International Conference on Learning Representations, ICLR 2021, Virtual Event, Austria, May 3-7, 2021*. OpenReview.net.
- [Shao et al., 2022] Shao, Z., Zhang, Z., Wang, F., and Xu, Y. (2022). Pre-training enhanced spatial-temporal graph neural network for multivariate time series forecasting. In *Proceedings of the 28th ACM SIGKDD conference on knowledge discovery and data mining*, pages 1567–1577.

- [Shi et al., 2015] Shi, X., 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. *Advances in neural information processing systems*, 28.
- [Strehl and Ghosh, 2002] Strehl, A. and Ghosh, J. (2002). Cluster ensembles—a knowledge reuse framework for combining multiple partitions. *Journal of machine learning research*, 3(Dec):583–617.
- [Sutskever et al., 2014] Sutskever, I., Vinyals, O., and Le, Q. V. (2014). Sequence to sequence learning with neural networks. In Ghahramani, Z., Welling, M., Cortes, C., Lawrence, N. D., and Weinberger, K. Q., editors, *Advances in Neural Information Processing Systems 27: Annual Conference on Neural Information Processing Systems 2014, December 8-13 2014, Montreal, Quebec, Canada*, pages 3104–3112.
- [Transportation Research Center (TRC) and Department of Civil and Coastal, 2007] Transportation Research Center (TRC) and Department of Civil and Coastal (2007). Model task force meeting: Activity-based modeling from an academic perspective. Activity-based modeling presentation and discussions.
- [U.S. Department of Transportation, 2017] U.S. Department of Transportation, F. H. A. (2017). 2017 national household travel survey.
- [Vaswani, 2017] Vaswani, A. (2017). Attention is all you need. *Advances in Neural Information Processing Systems*.
- [Wang et al., 2021] Wang, J., Jiang, J., Jiang, W., Li, C., and Zhao, W. X. (2021). Libcity: An open library for traffic prediction. In *Proceedings of the 29th International Conference on Advances in Geographic Information Systems, SIGSPATIAL ’21*, page 145–148, New York, NY, USA. Association for Computing Machinery.
- [Wang et al., 2020] Wang, M.-X., Lee, W.-C., Fu, T.-Y., and Yu, G. (2020). On representation learning for road networks. *ACM Transactions on Intelligent Systems and Technology (TIST)*, 12(1):1–27.
- [Wang et al., 2022] Wang, X., Wu, X., and Liu, J. (2022). Optimization models of actuated control considering vehicle queuing for sustainable operation. *Sustainability*, 14(15):8998.
- [Wu et al., 2019] Wu, Z., Pan, S., Long, G., Jiang, J., and Zhang, C. (2019). Graph wavenet for deep spatial-temporal graph modeling. *Proceedings of the Twenty-Eighth International Joint Conference on Artificial Intelligence (IJCAI-19)*.
- [Yu et al., 2018] Yu, B., Yin, H., and Zhu, Z. (2018). Spatio-temporal graph convolutional networks: A deep learning framework for traffic forecasting. In Lang, J., editor, *Proceedings of the Twenty-Seventh International Joint Conference on Artificial Intelligence, IJCAI 2018, July 13-19, 2018, Stockholm, Sweden*, pages 3634–3640. ijcai.org.
- [Zhao et al., 2019] Zhao, L., Song, Y., Zhang, C., Liu, Y., Wang, P., Lin, T., Deng, M., and Li, H. (2019). T-gcn: A temporal graph convolutional network for traffic prediction. *IEEE transactions on intelligent transportation systems*, 21(9):3848–3858.
- [Zheng et al., 2020] Zheng, C., Fan, X., Wang, C., and Qi, J. (2020). Gman: A graph multi-attention network for traffic prediction. In *Proceedings of the AAAI conference on artificial intelligence*, volume 34, pages 1234–1241.

Acknowledgments in Korean

카이스트는 저에게 너무나 뜻깊은 모교입니다.

한국과학영재학교에 와서 공부하기가 싫고 과학 실험 보고서를 쓰는 게 싫었던 저는 컴퓨터를 가지고 노는 게 좋았습니다. 불성실한 학생이라 기억도 못 하시겠지만, 김호숙 교수님의 강의를 들으며 컴퓨터 시뮬레이션 프로그램을 만들어보며 컴퓨터과학에 재미를 붙이기 시작했습니다.

카이스트 학부에 진학하고는 후회 없는 삶을 살자는 마음가짐으로 도전할 수 있는 것들을 도전하면서 살았습니다. 특히 교내창업대회 이후 변호사와 사람을 연결하는 채팅 서비스를 만들겠다는 스타트업인 로앤컴퍼니에 입사하게 되었습니다. 당시 로앤컴퍼니 개발팀 이휘진 이사님, 이상후 카이스트 선배님의 조언과 도움을 받으며 언젠가는 개발자가 대우받고 재밌게 일할 수 있는 페이스북이나 마이크로소프트와 같은 IT 스타트업을 만들겠다고 꿈꾸었습니다.

카이스트 심화전공 과정을 거치며 한태숙 교수님께서 “너희들이 잘 먹고 잘사는 게 나라를 위하는 일이다”라고 하신 게 기억이 납니다. 가장 이기적으로 행동하는 게 결국 가장 이타적인 것이 다와 같은 이야기였을까요. 그때 미국 여름학교로 실리콘밸리의 구글, IBM, 인텔을 방문하며 언젠간 이런 곳에서 일하고 싶다고 생각했던 기억이 납니다.

학부 수업에서는 운영체제를 가르쳐주시던 송준화 교수님께서 “연구란 지금까지 한 내용을 정리하는 것뿐 아니라 미래의 비전을 제시해야 한다”고 말씀하신 게 가장 기억에 남습니다. 비록 석사·박사 과정을 같이 하지는 못했지만 매주 연구 미팅 준비를 할 때면 그 말씀이 기억에 남습니다. 그리고 카이스트 전산 학부에 몰입캠프를 도입하여 학생들에게 스타트업을 경험할 수 있도록 힘써주신 류석영 교수님과 장병규 의장님께 전산학부 학생들을 대표하여 감사드립니다. 또한 저의 석박사 과정에서 저의 연구 주제를 제안해주시고 5년간 안정적으로 투자를 해주신 이동만 교수님, 그리고 함께 연구 지도를 도와주신 안지선 선배님이자 교수님, 매주 흥미로운 토론으로 카페에서 대화의 시간을 써주셨던 박영준 박사님께 감사드립니다. 그리고 저를 항상 위로해 주고 응원해 준 가족과 조시아, 마지막 연구를 정리하며 저 자신이 사회를 위해 무엇을 할 수 있을지 깨닫게 해준 이예지 학생에게 깊은 감사를 전합니다.

이 단 한 편의 박사논문을 만들기까지 이토록 수많은 사람들의 사랑과 노력이 들어갔다고 하기에는 완벽하지 못한 부분이 많아 부끄러운 마음입니다. 또한 저에게 도움을 주신 분들에게 고마움을 말 이상으로 표현할 방법이 없어 송구스럽습니다.

제가 원하는 세상은 카이스트라는 권위가 필요 없는 세상입니다. 취업을 위한 학점이나 논문 개수가 아니라 진정으로 배움을 위해서 대학을 다닐 수 있는 세상을 만들고 싶습니다. 그렇기 때문에 더더욱 권위라는 물질적 가치를 벌리지 않고 인간의 순수한 직관으로 타인의 가치를 알아볼 수 있는 문화를 만들고 싶습니다.

아이러니하게도 그러한 세상을 만들기 위해 저는 반대의 행보를 걷고 있습니다. 카이스트라는 대학을 중퇴하더라도 실력만으로 인정받고 싶었던 젊은 날의 수민이를 위해서, 그러한 사람을 진심으로 응원해 주고 불안해하지 않도록 하는 든든한 서포터가 되기 위해 박사의 길을 택했던 것 같습니다.

이 논문은 제 이야기의 마무리가 아닌, 시작점으로, 제가 누렸던 교육의 기회를 누리지 못한 수많은 사람들을 위해, 그리고 꿈과 재능이 있지만 학위라는 권위에 휘둘려 사회적 억압을 받으며 젊음을 낭비해야 했던 수많은 인재가 더 이상 방황하지 않도록 평생 봉사하는 마음으로 살아가고자 합니다. 누군가는 해야 하지만 누구나 하지 못하는 일을 하며 살겠습니다.

한국의 경제성장을 위해 카이스트를 설립하는 데 공헌하신 실리콘밸리의 아버지 터만 박사님, 그리고 진리를 찾아 방황할 때 항상 영감을 주셨던 저의 영원한 스승, 쇼펜하우어 님께 이 박사논문을 바칩니다.

2024년 12월 18일
대전에서

Curriculum Vitae

Name : Sumin Han

Date of Birth : December 18, 1994

E-mail : suminkaist@gmail.com

Homepage : <https://smhanlab.com/>

Educations

- 2018. 3. – 2025. 2. KAIST (Integrated MS. and Ph.D. in Computer Science)
- 2015. 6. – 2015. 7. University of California, Santa Cruz (Summer School)
- 2013. 3. – 2018. 2. KAIST (Bachelor of Science in Computer Science)
- 2011. 6. – 2011. 7. University of Toronto (Summer Camp)
- 2009. 3. – 2013. 2. Korea Science Academy of KAIST

Career

- 2017. 7. – 2017. 8. Satrec Initiative Co., Ltd. (Software Engineering Intern) – Parallel Computing Framework Design (Docker, Slurm, Ganglia)
- 2014. 3. – 2014. 5. Law&Company Co., Ltd. (Software Engineer) – Web Development (Node.js, MongoDB, HTML, CSS)

Publications

1. **Sumin Han**, Youngjun Park, Minji Lee, Jisun An, and Dongman Lee. “Enhancing Spatiotemporal Traffic Prediction through Urban Human Activity Analysis.” *Proceedings of the 32nd ACM International Conference on Information and Knowledge Management (CIKM '23)*, 2023.
2. **Sumin Han**, Jisun An, and Dongman Lee. “Spatio-Temporal Road Traffic Prediction using Real-time Regional Knowledge.” *ArXiv*, 2022.
3. **Sumin Han**, Jisun An, Youngjun Park, Suji Kim, Kitae Jang, and Dongman Lee. “Multiple Area Feature-aware Spatio-temporal Public Transit Prediction.” *ArXiv*, 2021.
4. Youngjun Park and **Sumin Han**. “Enhancing Population Predictions in Developing Area through Building Construction Data.” *Association of European Schools of Planning (AESOP)*, 2024.
5. Youngjun Park, **Sumin Han**, Jisun An, and Dongman Lee. “Encoding Urban Trajectory as a Sentence: Deep Learning Insights for Human Mobility Pattern.” *ACSP*, 2023.
6. **Sumin Han**, Dasom Hong, and Dongman Lee. “Exploring Commercial Gentrification using Instagram Data.” *2020 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining (ASONAM)*. IEEE, 2020, pp. 557–564.

7. Sumin Han, Kinam Park, and Dongman Lee. “Discovering Daily POI Exploitation Using LTE Cell Tower Access Traces in Urban Environment.” *International Conference on Social Informatics (SocInfo)*. Springer, 2020, pp. 81–94.
8. Dongmin Kim, Sumin Han, Heesuk Son, and Dongman Lee. “Human Activity Recognition Using Semi-supervised Multi-modal DEC for Instagram Data.” *Advances in Knowledge Discovery and Data Mining: 24th Pacific-Asia Conference, PAKDD 2020*, Singapore, May 11–14, 2020, Proceedings, Part I. Springer, 2020, pp. 869–880.
9. (국내) Ahmad Elmoursi, Sumin Han, Hyungchan Bae, Youngjun Park, Nguyen Thanh Tung, and Dongman Lee. “Enhancing Geospatial Recommendations through Natural Language Processing with Large Language Models.” *한국정보과학회 학술발표논문집*, 2024.
10. (국내) Youngjun Park, Sumin Han, Doyoon Park, Hyeongchan Bae, Nguyen Thanh Tung, and Dongman Lee. “Integrative Geospatial Visualization for Urban Mobility Analysis.” *한국정보과학회 학술발표논문집*, 2024.
11. (국내) Thanh Long Cao, Sumin Han, and Dongman Lee. “Applying Scene Graphs for Improving Multi-modal Supervised Learning based Human Activity Recognition on Instagram Posts.” *한국정보과학회 학술발표논문집*, 2022, pp. 855–857.
12. (국내) Changhui Kim, Sumin Han, and Dongman Lee. “Inference of Station-unit Placeness through a Subway Ridership Pattern.” *한국정보과학회 학술발표논문집*, 2022, pp. 1794–1796.
13. (국내) Zofia Marciniak, Sumin Han, and Dongman Lee. “Understanding the Impact of the Weather on Human Mobility via LTE Access Traces in Seoul Districts.” *한국 HCI 학회 학술대회*, 2021, pp. 615–618.
14. (국내) 김수빈, 한수민, and 이동만. “서울시의 토지 이용에 따른 도로의 차량 속도의 상관관계 분석.” *한국정보과학회 학술발표논문집*, 2021, pp. 119–121.
15. (국내) 안정미, 한수민, and 이동만. “인스타그램 데이터를 활용한 장소성 분석 및 시각화.” *한국 HCI 학회 학술대회*, 2020, pp. 929–933.
16. (국내) 정종구, 한수민, and 이동만. “인스타그램에서 나타나는 인간 활동 패턴의 연간 변화 분석.” *한국 HCI 학회 학술대회*, 2020, pp. 723–726.
17. (국내) 한수민 and 김명호. “베이지안 개인화 순위 방법 기반 대용량 이종 네트워크에서의 효율적인 랜덤 워크 순위 계산.” *2018 한국소프트웨어종합학술대회*. 한국정보과학회, 2018.

“THE World is My Idea”

Arthur Schopenhauer (1788 – 1860)