National Research University Higher School of Economics

as a manuscript

Seungmin Jin

A VISUAL ANALYTICS SYSTEM FOR EXPLAINING AND IMPROVING ATTENTION-BASED TRAFFIC FORECASTING MODELS

PhD Dissertation Summary

for the purpose of obtaining academic degree
Doctor of Philosophy in Computer Science

Academic supervisor:
Doctor of Sciences
Attila Kertész-Farkas

Moscow — 2024

Contents

1 Introduction

I investigate the error of spatio-temporal graph attention networks model (ST-GRAT) [1] in the traffic forecasting problem through my visual analytics system. My goal is to improve the model's performance based on the understanding of hidden state and patterns of error [2]. Ultimately, it leads to more accurate traffic prediction and enables effective development for the deep traffic forecasting models. In this thesis, I explain what ST-GRAT is, then demonstrate the analysis of the model's prediction and error, debugging process of the model using my visual anlaytics system, called AttnAnalyzer [3].

Main contributions of my dissertation as follow:

1. I introduce the spatio-temporal graph attention networks model (ST-GRAT) [1] that I developed to predict the traffic (Chpt.2). This chapter also provides the literature review of traffic forecasting problem, the definition using graph neural networks including temporal and spatial dependency, and the explanation of attention mechanism to decompose the prediction.

2. I explain the visual analytics system to understand the prediction of ST-GRAT in traffic forecasting (Chpt.3). This chapter introduce research requirements of the visual analytics system, automated methodology and its implementation to understand model's hidden space and prediction patterns.

3. I show the analysis of ST-GRAT model using two different road networks, the urban roads and the highway. This chapter explains the problem of the model and show the method how to improve the performance by hacking the attention networks through my visual analytics system, called AttnAnalyzer [3]. (Chpt.4).

4. I put down the concussion of my disetration and future works (Chpt.5).

1.1 Background and Motivation

Traffic congestion has become a significant issue in urban areas worldwide, resulting in increased travel times, environmental pollution, and reduced economic productivity [4, 5]. Accurate traffic forecasting is essential to mitigate these issues and provide better traffic management solutions. Spatio-temporal graph attention networks have emerged as a promising approach for traffic forecasting, owing to their ability to capture complex spatial and temporal dependencies in traffic data [5]. Among these models, the ST-GRAT model has shown promising results in predicting dynamically changing road speeds [1].

Although attention-based models show good performance, there are issues to deploy this type of model in the real world. The biggest issue is the reliability of model's prediction, since we have the report of average performance but don't understand the behavior of the model. The behavior indicates the pattern of error in the prediction. For example, we don't know when and in which case these model fails, furthermore why these kind of error happens. In average, the model may provide good prediction, but if it keeps showing wrong forecasting, the credit issue would arise eventually from users [6].

Another serious issue is the high cost of debugging. The graph attention networks that represents the spatio-temporal dependency is huge and complex [7, 1], it is really difficult and time-consuming to debug the model. Furthermore, the deep learning model itself requires a lot of training time for the debugging process [8]. This thesis challenges these issues, or limitations, by understanding the behavior of the model and developing methods to improve the debugging process.

1.2 The relevance of research

To address these limitations, it is crucial to understand the behavior of deep learning models in traffic forecasting and to develop methods to improve their performance, such like understanding the hidden space for the deep learning classification problem [9]. This research is highly relevant in light of the increasing demand for accurate traffic prediction in urban areas, where traffic congestion is a major problem. Furthermore, the development of more effective deep learning models for traffic forecasting can have broader implications for transportation management and urban planning [6].

Several recent studies have also emphasized the need for a deeper understanding of the behaviors of deep learning models for traffic forecasting [5]. For example, in a study by Zhang et al. (2020) [10], the authors proposed a novel attention mechanism to improve the performance of traffic prediction models. In another study by Huang et al. (2020) [11], the authors used a multimodal deep learning approach to predict traffic flow and congestion. These studies demonstrate the importance of developing more effective deep learning models for traffic forecasting, as well as the potential benefits of gaining a deeper understanding of these models' behaviors.

1.3 Research Objectives and Scope

This thesis aims to enhance the performance of the ST-GRAT model, which is the state-of-art attention-based deep learning model, by incorporating a visual analytics system to understand the reasons behind the model's failure and address them. The research objectives are:

1. Investigate the factors causing the ST-GRAT model's failure in accurately forecasting traffic.
2. Develop a visual analytics system to understand and analyze the model's behavior and shortcomings.
3. Improve the ST-GRAT model based on the insights gained from the visual analytics system.
4. Evaluate the enhanced model's performance in traffic forecasting and its potential impact on traffic congestion management.

The scope of this research covers the analysis of the ST-GRAT model, the development of a visual analytics system, and the improvement and evaluation of the model's performance.

1.4 Importance of work

The main contributions of this work include the following: 1) the deep traffic forecasting model using graph attention networks, 2) a visual analytics (VA) system design for exploring traffic forecasting model's pattern of prediction from a spatio-temporal perspective, 3) incorporation of automated methods, Dynamic Time Warping (DTW) [12], Granger causality test [13] and clustering for visual temporal analysis, 4) development of an attention enforcement method, 5) quantitative and qualitative evaluations of the system with three case studies to demonstrate how to explain deep learning models with attention, proven model's accuracy improvements with the attention enforcement method, and domain experts' feedback, and 6) we show how model designers can improve model performance using our tool by developing improved version of model based on the findings in the case study.

To my knowledge, this work is the first attempt exploring the attention-based traffic forecasting models' prediction process. This work also firstly tries to improving performance in the traffic domain by demonstrating the power of visual analytics approaches [14]. Traffic data is heterogeneous with extreme cases [6, 4, 15] and affected by uncontrollable external factors, such as accidents. Thus the traffic prediction task is especially challenging in that the models in the domain need to learn not only spatio-temporal features from the data, but also how to respond to implicit external events on roads. The external factors could even vary by region [7], which further challenges the models.

1.5 Publications

My thesis is based on following 3 main research papers, all of them have been published in Q1 journals or A* Conferences. Ranking is based on Scopus and Web of Science.

First-tier publications.

1. Seungmin Jin, Hyunwook Lee, Cheonbok Park, Hyeshin Chu, Yunwon Tae, Jaegul Choo, Sungahn Ko. "A visual analytics system for improving attention-based traffic forecasting models." **IEEE transactions on visualization and computer graphics, Q1 journal**, 2022, doi: https://doi.org/10.1109/tvcg.2022.3209462.

   Main Contribution: As a main author, I designed the whole research process, the visual analytics system and the performance improving method of ST-GRAT. I not only discovered the pattern of errors of ST-GRAT through detail case stduies, but also showed that how we can fix the error using representative speed patterns.

2. Cheonbok Park, Chunggi Lee, Hyojin Bahng, Yunwon Tae, Kihwan Kim, Seungmin Jin, Sungahn Ko and Jaegul Choo. "ST-GRAT: A novel spatio-temporal graph attention networks for accurately forecasting dynamically changing road speed." **Proceedings of the 29th ACM international conference on information & knowledge management (CIKM), ACONF**, 2020, doi: https://doi.org/10.1145/3340531.3411940.

Main Contribution: : I participated as the one of the main model developers and the data analyst. Especially, I debugged and analyzed a lot of this model to understand in which cases ST-GRAT predicts good or bad. In the research I could find the motivation why I need to develop the visual analytics system to understand behaviors of the deep learning models. I've experienced that the process of debugging is not only difficult to analyze, but also consumes a lot efforts and time because of the complexity.

3. Chunggi Lee, Yeonjun Kim, Seungmin Jin, Dongmin Kim, Ross Maciejewski, David Ebert, Sungahn Ko. "A visual analytics system for exploring, monitoring, and forecasting road traffic congestion." **IEEE transactions on visualization and computer graphics, Q1 journal**, 2020, doi: https://doi.org/10.1109/tvcg.2019.2922597.

   Main Contribution: I participated as the one of the main model developers and the data analyst. In this research, I developed the whole visual analytics systems that shows traffic congestion for the city of Ulsan, South Korea. I also performed several cases studies that analyze in which conditions traffic jam happens. I also found the motivation of my primary research, since the deep learning model, LSTM, I used here does not provide structured information of inference, so it was difficult to understand.

Other publications.

Although all following papers published in Q1 or journals or A* Conferences, they are not the basis in this thesis.

1. Hyeshin Chu, Joohee Kim, Seongouk Kim, Hongkyu Lim, Hyunwook Lee, Seungmin Jin, Jongeun Lee, Taehwan Kim, and Sungahn Ko. "An Empirical Study on How People Perceive AI-Generated Music." **Proceedings of the 31st ACM International Conference on Information & Knowledge Management (CIKM), ACONF, 2022. doi:** https://doi.org/10.1145/3511808.3557235

2. Beknazarov, Nazar, Seungmin Jin, and Maria Poptsova. "Deep learning approach for predicting functional Z-DNA regions using omics data." **Scientific Reports 10.1 (2020): 19134., Q1 journal. doi:** https://doi.org/10.1038/s41598-020-76203-1.

3. Hyunwook Lee, Seungmin Jin, Hyeshin Chu, Hongkyu Lim, and Sungahn Ko."Learning to Remember Patterns: Pattern Matching Memory Networks for Traffic Forecasting" **International Conference on Learning Representations (ICLR), ACONF**, 2022, doi: https://doi.org/10.48550/arXiv.2110.10380.

4. Hyunwook Lee, Cheonbok Park, Seungmin Jin, Hyeshin Chu, Jaegul Choo, and Sungahn Ko."An Empirical Experiment on Deep Learning Models for Predicting Traffic Data" **IEEE 37th International Conference on Data Engineering (ICDE), ACONF**, 2021, doi: https://doi.org/10.1109/icde51399.2021.00160.

2 Explanation of Spatio-Temporal Graph Attention Networks (ST-GRAT) Model and Attention Mechanism

Traffic forecasting plays a crucial role in predicting future traffic conditions based on historical data [4]. In recent years, deep learning approaches have gained significant attention for their ability to capture complex patterns and dependencies in traffic data. However, in my previous, I notice that understanding the behavior of deep learning models is very difficult since its complexity of spatio temporal dependecies [7]. This is my main motivation starting to develop ST-GRAT by utilizing attention mechanisms, since it can provide intuitive information of hidden space as references. This chapter explores important concepts and techniques related to traffic forecasting using deep learning, including attention mechanisms, graph neural networks (GNNs), Dynamic Time Warping (DTW), Granger causality, and Spatio-Temporal Graph Attention Networks (ST-GRAT) [1] which is the model I have analyzed.

2.1 Attention Mechanism

Attention mechanisms have emerged as a powerful technique in deep learning and neural networks [16, 4]. They allow models to focus on specific parts of the input data when making predictions or decisions, enabling them to assign different weights or importance to different elements based on their relevance to the task at hand [17].

Formally, an attention mechanism computes a weighted sum of value vectors with size of d based on the similarity or compatibility between a query vector $q \in \mathbb{R}^d$, a set of key vectors $K = k_1, k_2, ..., k_n \in \mathbb{R}^d$, and a set of value vectors $V = v_1, v_2, ..., v_n \in \mathbb{R}^d$. The attention mechanism h_{att} can be represented as:

$$\text{Attention}(q, K, V) = h_{\text{att}}(q, K, V) = \sum_{i=1}^{n} \alpha_i v_i,$$

where α_i is the weight assigned to the i-th value vector. The weights are calculated based on a compatibility function $f(q, k_i)$, measuring the similarity or relationship between the query vector q and the i-th key vector k_i. Typically, the weights α_i are computed using a softmax function over the compatibility scores, ensuring they sum up to 1.

The attention mechanism is widely used in various domains, including natural language processing, computer vision, and time series analysis. It has been employed for tasks such as machine translation, image captioning, and sentiment analysis [16].

In the context of traffic forecasting, attention mechanisms can be incorporated into deep learning architectures to capture relationships and dependencies within traffic data. By attending to different parts of the input data with varying importance, models can effectively extract relevant information and improve their performance and interpretability.

2.2 Traffic Forecasting Problem in Graph Neural Networks

Traffic forecasting aims to predict future traffic conditions based on historical data. In the context of graph-based traffic forecasting, the traffic network is represented as a graph, where the nodes represent road segments and the edges represent the connectivity between them. Each node in the graph is associated with a traffic attribute, such as speed or volume, types of road, the number of lanes and etc [1].

Graph neural networks (GNNs) have emerged as a powerful approach for traffic forecasting in graph-structured data [18]. GNNs leverage the inherent spatial and temporal dependencies in the traffic network to capture complex patterns and make accurate predictions. In a GNN, each node represents a traffic sensor, and the edges represent the correlations between sensors. By propagating information between nodes and aggregating local features from neighbors, GNNs can learn the spatial dependencies in the traffic flow observed at different locations. By processing sequential traffic data, GNNs can also learn the temporal dependencies in the traffic flow observed at different time periods [19].

One key component in GNNs is the attention mechanism, which allows the model to focus on important features or nodes in the traffic network. The attention mechanism in GNNs learns to assign different weights or importance to the nodes in the traffic network based on their relevance to the prediction task. By attending to relevant nodes, the model can capture the spatial dependencies and incorporate them into the forecasting process [20]. The attention mechanism enables the model to adaptively allocate more attention to informative nodes while downplaying the influence of less informative nodes [1].

In the context of traffic forecasting, the attention mechanism can learn to attend to neighboring nodes that have a strong influence on the target node's traffic behavior. This is achieved by computing attention weights based on the compatibility or similarity between the query node and the key nodes in the graph[1, 17]. The attention weights are then used to compute a weighted sum of the value nodes, which represent the features or attributes of the neighboring nodes.

A weighted directed Graph G, which represents the underlying road network, forms the cornerstone of this research. It is defined as $G = (V, E, A)$, where:

- V denotes the set of nodes, with $|V| = N$.
- E stands for the set of edges.
- $A \in \mathbb{R}^{N \times N}$ is a weighted adjacency matrix that encodes the proximity between nodes, often derived from road network distances.

The adjacency matrix A encapsulates two pivotal aspects of proximity within the road network:

1. Connectivity: This element of A indicates whether two nodes are directly connected or not.
2. Edge Weights: Edge weights in A provide a more nuanced understanding. They not only acknowledge the presence of a connection but also encompass finer details like the distance between connected nodes and the direction of the connection.

Collectively, these indicators of proximity shape the overall structure of the graph, encompassing characteristics like connectivity, edge orientations, and inter-node distances.

We now dive into graph signal X, which is our matrix representation of the graph G. Graph signal X, represents the observed traffic flow, and it is denoted as $X \in \mathbb{R}^{N \times P}$. Here, P represents the number of features associated with each node. These features can be a range of attributes including velocity, volume, road types, and lane counts. However, note that ST-GRAT uses only the information of velocity as a feature and forecasts the future speed.

The interaction between G and X holds significance:

- The nodes of graph G correspond to rows within matrix X.
- Each row in X captures the specific features of traffic flow pertaining to the corresponding node.

This symbiotic relationship between G and X lays the foundation for analyzing traffic dynamics and patterns within the road network. In subsequent discussions, we will explore how this relationship forms the basis for traffic forecasting, augmented by the integration of an attention mechanism for refined predictions.

Let $X(t) \in \mathbb{R}^{N \times P}$ represent the graph signal observed at time t, the traffic forecasting problem aims to learn a function h that maps T' historical graph signals to future T graph signals, given a graph G:

$$[X(t - T^{'} + 1), \cdots, X(t); G] \xrightarrow{h} [X(t + 1), \cdots, X(t + T)]. \tag{1}$$

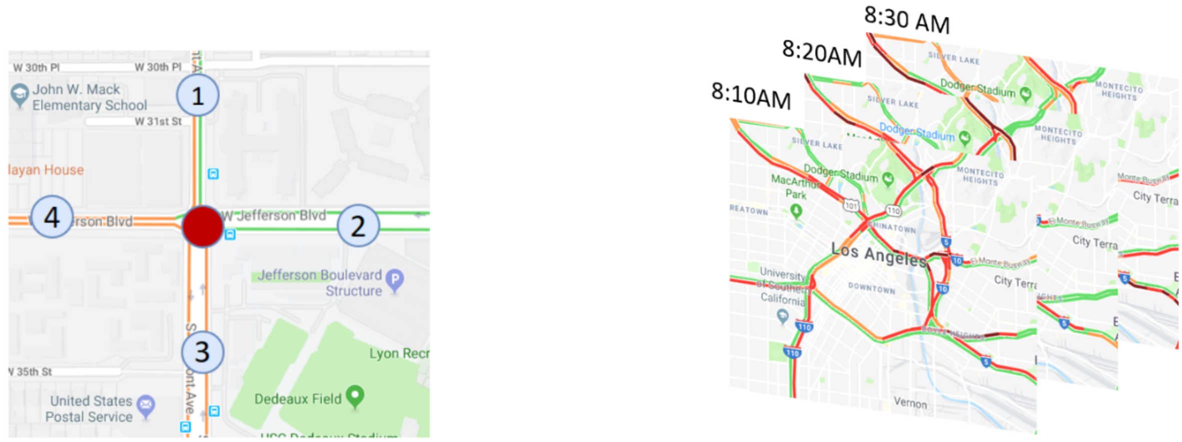
This formula represents the traffic forecasting problem, where the goal is to predict future traffic speed given previously observed traffic flow from N correlated sensors on the road network. The attention mechanism is embedded within the function h, which is responsible for learning the mapping from the input sequence $[X(t - T^{'} + 1), \cdots, X(t); G]$ to the output sequence $[X(t + 1), \cdots, X(t + T)]$. The attention mechanism $h_{\text{att}}(q, K, V)$ within h allows the model to assign different weights or importance to different elements (e.g., nodes, features) in the input sequence based on their relevance to the prediction task. By doing so, the model can dynamically focus on the most informative parts of the data during the forecasting process.

GNNs with attention mechanism have shown great potential in traffic forecasting by effectively capturing both spatial and temporal dependencies in traffic flow. With the increasing availability of traffic data and the development of new GNN models, I expect GNNs with the attention mechanism to become an increasingly popular method for traffic forecasting [20, 18, 5].

2.3 Spatio-Temporal Graph Attention (ST-GRAT)

I demonstrate the utilization of ST-GRAT in showcasing my visual analytics (VA) approach. I specifically choose ST-GRAT for two reasons: 1) it has demonstrated state-of-the-art performance, and 2) it allows me to generate attention matrices to analyze spatio-temporal dependencies.

Figure 3 provides an overview of my system workflow, which consists of various functional modules and the capability to answer questions based on the data. The workflow includes data



(a) Which neighbor road has the largest effects?

(b) Which historical observation is more relevant?

Figure 1: Interpretable traffic forecasting. The spatial dependency (a) and temporal dependency (b) are the most important information in the traffic prediction [6].

pre-processing (A), automated methods for supporting spatio-temporal analysis (B), model training and inference (C), visualization modules (D), and the ability to answer questions using the system (E).

ST-GRAT is based on the transformer architecture and follows an encoder-decoder structure with self-attention (i.e., temporal attention). In addition, it incorporates graph attention as the spatial attention mechanism before the temporal attention, using a sentinel vector for skip connections within the same road. ST-GRAT utilizes 12-length sequential historical speed data with encoded features for each road and predicts 12 sequential speed predictions.

The overall architecture of ST-GRAT is depicted in Figure 2. It comprises three types of layers: embedding, spatial attention, and temporal attention layers. The embedding layer takes the road network, speed, and observed time as input features and utilizes the position embedding method to encode the order of the given sequence. The spatial attention layer captures spatial dependencies among neighboring roads through a graph attention network, improving dependency modeling and interpretability. The temporal attention layer models the temporal dependency and trends of the given sequences by performing multi-head attention.

To analyze the spatio-temporal dependencies, I derive the spatio-temporal attention matrix (ST matrix) from each attention layer. This matrix represents the relationship between the input sequence before spatial attention and the output sequence after temporal attention. Specifically, for each attention head, the relationship can be written as follows:

$$H = (TA \odot SA)X,$$

where $TA \in \mathbb{R}^{\mathbb{N}\times\mathbb{T}}$ and $SA \in \mathbb{R}^{\mathbb{N}\times\mathbb{T}}$. Note that ST-GRAT uses $\mathrm{T} = 12$ for the sequences of window. $TA \odot SA$ represents the spatio-temporal attention, which is the attention of interest for understanding model behaviors.

In summary, ST-GRAT is a variation of the transformer model that combines spatial and temporal attention mechanisms to capture spatio-temporal dependencies. It leverages the strengths

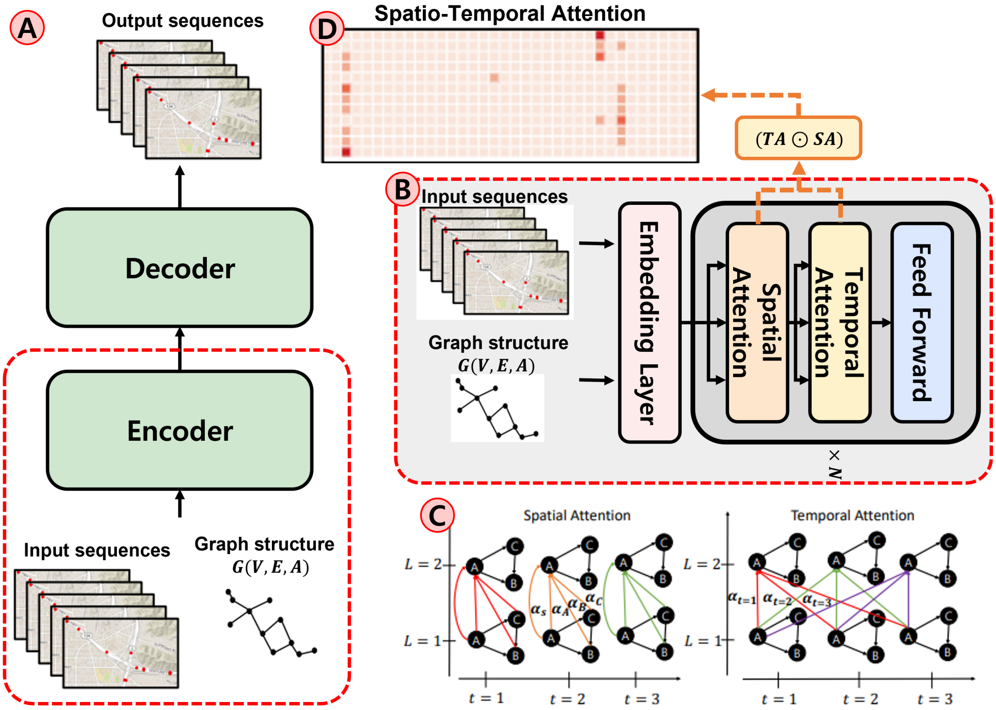


Figure 2: Overall architecture of ST-GRAT (A). Each layer in the encoder is composed of a stack of identical layers (B). G(V, E, A) in (A) indicates a set of nodes, links, and a weighted adjacency matrix, respectively. I create a spatio-temporal attention matrix (D) by calculating point-wise multiplication between spatial and temporal attention weights (C).

Table 1: ST-GRAT & PM-MemNet Prediction accuracy on Los Angeles Dataset

	T	Metric	GCRNN	DCRNN	GaAN	STGCN	Graph WaveNet	HyperST	GMAN	ST-GRAT	PM-MemNet
METR-LA	15 min	MAE	2.80	2.73	2.71	2.88	2.69	2.71	2.81	**2.60**	2.65
		RMSE	5.51	5.27	5.24	5.74	5.15	5.23	5.55	**5.07**	5.29
		MAPE	7.5%	7.12%	6.99%	7.62%	6.90%	-	7.43%	**6.61**%	7.01%
	30 min	MAE	3.24	3.13	3.12	3.47	3.07	3.12	3.12	**3.01**	3.03
		RMSE	6.74	6.40	6.36	7.24	6.26	6.38	6.46	**6.21**	6.29
		MAPE	9.0%	8.65%	8.56%	9.57%	8.37%	-	8.35%	**8.15** %	8.42%
	1 hour	MAE	3.81	3.58	3.64	4.59	3.53	3.58	**3.46**	3.49	**3.46**
		RMSE	8.16	7.60	7.65	9.40	**7.37**	7.56	**7.37**	7.42	**7.29**
		MAPE	10.9%	10.43%	10.62%	12.70%	**10.01%**	-	10.06%	**10.01%**	**9.97%**
	Average	MAE	3.28	3.14	3.16	3.64	3.09	3.13	3.13	**3.03**	**2.99**
		RMSE	6.80	6.42	6.41	7.46	6.26	6.39	6.46	**6.23**	**6.14**
		MAPE	9.13%	8.73%	8.72%	9.96%	8.42%	-	8.61%	**8.25%**	8.27%

of attention models in various tasks, including speed prediction, travel time estimation, and taxi demand prediction [16, 74]. Table 1 shows that ST-GRAT outperforms other GNNs based models. The attention matrices generated by ST-GRAT facilitate the analysis of spatio-temporal relationships and enhance the interpretability of the model.

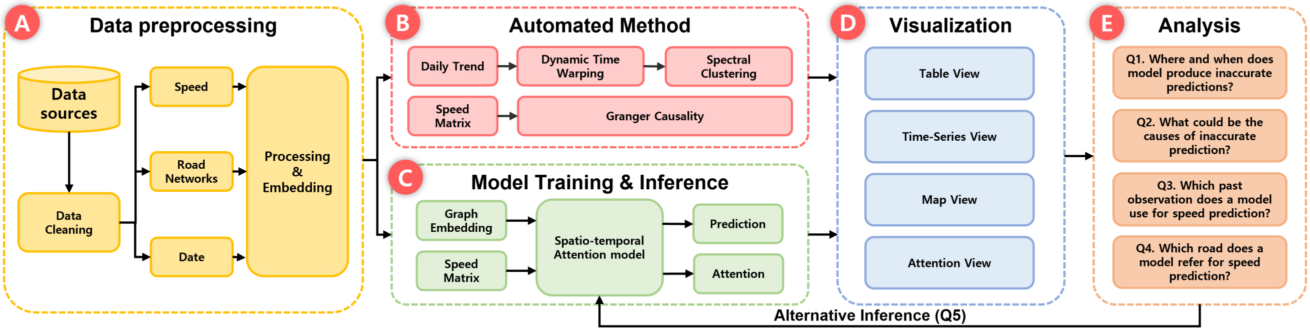


Figure 3: An overview of the system workflow with functional modules and questions. (A) Data pre-processing, (B) automated methods for supporting spatio-temporal analysis, (C) model training and inference, (D) visualization modules, and (E) answering questions using the system.

3 Introduction of Visual Analytics System of Traffic Forecasting, AttnAnalyzer

Visual analytics techniques have gained popularity in various domains, including traffic analysis. These techniques leverage interactive visualizations to help analysts understand and explore complex data sets. In the context of traffic analysis, visual analytics can aid in understanding spatio-temporal patterns of traffic flows, identifying factors contributing to congestion and traffic accidents, and supporting decision-making processes [21].

Visual analytics systems provide interactive tools and visualizations that enable analysts to explore and analyze traffic data in an intuitive and efficient manner. By visualizing spatial and temporal patterns of traffic flows, analysts can detect trends, identify anomalies, and gain insights into traffic behavior. These systems often incorporate machine learning algorithms and data exploration techniques to facilitate the exploration and interpretation of traffic data [21, 7, 6].

In this section, I present the system design and methodology that I have developed to address the task requirements derived from discussions with domain experts. The main objective of my system, called **AttnAnalyzer**, is to provide a visual analytics platform for exploring the behavior and performance of traffic forecasting models.

3.1 Task Description and Requirements

To design my system, I collaborated with three domain experts from a corporate organization that handles a significant volume of map and navigation services. These experts provided valuable insights into the challenges they face in optimizing navigation services and the limitations of existing machine learning models. Through extensive discussions over 18 months, I extracted the following task requirements:

- **Requirement R1:** The system should highlight roads with low accuracy and provide information on when the model produces inaccurate predictions.
- **Requirement R2:** The system should provide effective methods for exploring the encoded

spatio-temporal dependencies among roads to identify the causes of inaccurate predictions and understand which past observations the model utilizes for speed prediction.

- **Requirement R3:** The system should support users in formulating and validating hypotheses for improving model accuracy and provide information on how much improvement can be expected.

These requirements were formulated based on the primary questions posed by the experts, such as identifying inaccurate predictions, understanding the causes of inaccuracies, and improving model performance across different road topologies. Additionally, following questions were considered to provide better insights into the model's behavior and dependencies.

- **Q1: At which locations and times does the model generate inaccurate predictions?**
- **Q2: What potential factors contribute to inaccuracies in predictions?**
- **Q3: What historical (combinations of) observations does the model utilize for speed prediction?**
- **Q4: Which specific road does the model turn to for speed prediction?**
- **Q5: What strategies can be employed to enhance the model's accuracy?**
- **Q6: How does the model operate across varying road configurations?**

Among these inquiries, Q1 and Q2 assume primary importance, while Q3 and Q4 serve as ancillary inquiries that require clarification to better address Q1 and Q2. Q5 can be regarded as a question aimed at boosting market competitiveness, and Q6 represents the final question requiring attention before real-world deployment.

Based on the deliberations, we have distilled the ensuing prerequisites for a visual analytics (VA) system tailored to exploring the behavior of traffic forecasting models and enhancing their performance. Initially, since it is vital to identify problematic roads associated with a given model, **(R1) a VA system should spotlight roads exhibiting subpar accuracy and furnish insights into instances when the model's accuracy is compromised (Q1)**. Given that models predict forthcoming speeds through spatio-temporal interdependencies among roads, **(R2) a system should furnish an approach for effectively investigating these encoded dependencies (Q2–Q4)**, allowing users to uncover evidence pertaining to the correlation between elevated errors and speed patterns [22, 6]. Example data supporting the exploration of spatio-temporal dependencies encompass **(R2-1) historical traffic trends of roads, speed distribution patterns, standard deviation, daily speed trends (Q3, Q4)**, and **(R2-2) insights into model behavior (Q2–Q5)**, including the roads influencing a model's prediction (i.e., the roads impacting forecasts for a specific target road [23]) and the critical input sequences for predictions. Identifying roads influencing the prediction performance of a given road has often posed challenges for users [6]. Although prior research has demonstrated [23] the strong interdependency among neighboring roads and those connected secondarily and tertiarily through cross-validation, methodologies for determining whether a target road references suitable neighboring roads have been sparse. Consequently, **a system should also furnish insights into the similarity of temporal data and causal relationships among roads (Q2–Q4).** Lastly, to facilitate users' formulation and validation of hypotheses, **(R3) a system should provide an approach showcasing the potential improvement users can anticipate (Q5, Q6).**

By integrating the derived task requirements and research questions into my system design, I aim to provide a comprehensive solution that empowers domain experts to gain deeper insights into the behavior and performance of traffic forecasting models.

3.2 Integration of Dynamic Time Warping (DTW) and Granger Causality Tests

To analyze the spatio-temporal dependencies in traffic data, AttnAnalyzer integrates the use of Dynamic Time Warping (DTW) and Granger causality tests.

1. **Dynamic Time Warping (DTW):** In the context of traffic analysis, DTW can be used to compare the similarity between two traffic time series from different locations or different time periods. AttnAnalyzer utilizes DTW to measure the similarity between traffic patterns in different road segments. By clustering the time series based on their DTW distances, analysts can identify groups of segments with similar traffic patterns and gain insights into the temporal dependencies and trends in traffic flow.

2. **Granger Causality:** In the context of traffic analysis, Granger causality can be used to identify the causal relationship between different sensors in the road network. AttnAnalyzer utilizes Granger causality tests to analyze the dependencies between different road segments and identify the factors that contribute to congestion and traffic accidents. By examining the Granger causality between different pairs of sensors, analysts can gain insights into the flow of traffic in the network and identify potential bottlenecks or areas of congestion.

3.3 Overview of the Different Views and Visualizations Provided by AttnAnalyzer

AttnAnalyzer offers a variety of views and visualizations to support data exploration, feature selection, and model interpretation. These include:

1. **Map View:** The map view presents the road network and visualizes traffic attributes such as speed, volume, and congestion levels. Users can interact with the map view to select specific road segments, zoom in and out, and explore the spatial distribution of traffic patterns. Different colors or icons can represent different traffic attributes, allowing users to identify areas of interest and observe changes over time.

Table 2: Improved accuracy using the findings from my visual analytics approach. 10% of roads with the highest error are used.

Dataset	Method	15 Mins	30 Mins	60 Mins	Average
Ulsan	Graph WaveNet	9.57	9.87	10.52	9.98
	ST-GRAT	8.70	9.00	9.52	8.98
	Enforced w/ DTW	**8.23**	8.43	8.81	8.39
	Enforced w/ DTW + Granger	**8.24**	**8.37**	**8.76**	**8.37**
METR-LA	Graph WaveNet	6.68	8.17	11.66	8.83
	ST-GRAT	6.08	7.43	10.61	7.68
	Enforced w/ DTW	6.06	7.38	10.24	7.55
	Enforced w/ DTW + Granger	**6.04**	**7.28**	**10.08**	**7.48**

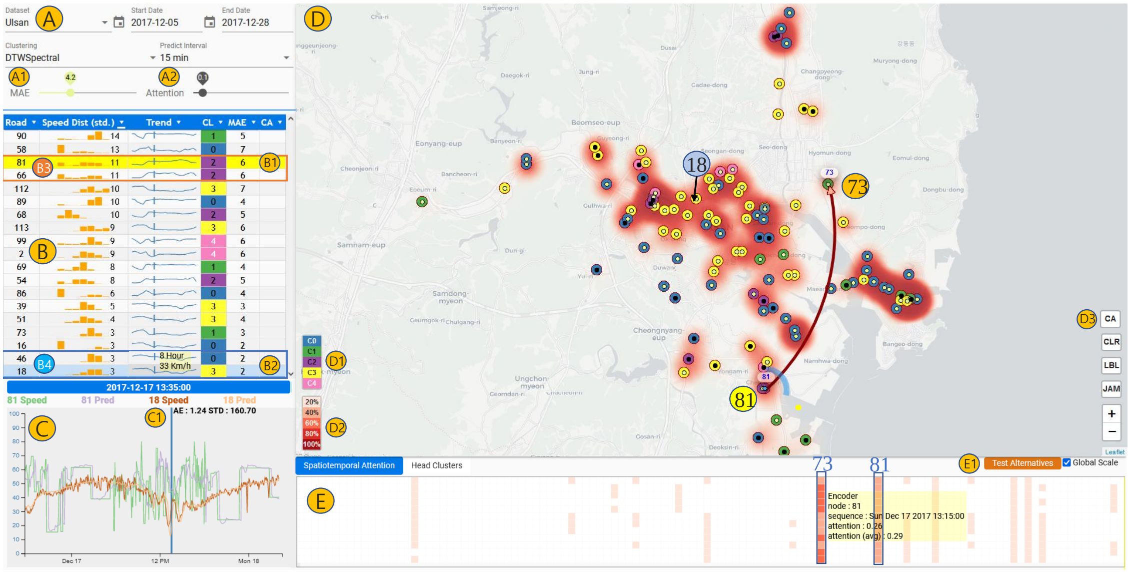


Figure 4: **AttnAnalyzer:** My visual analytics system for analyzing how an attention-based deep learning model predicts traffic congestion. (a) filter view, (b) table view, (c) ground-truth&prediction result comparison view, (d) a map with attention curves and clusters, and (e) attention heatmap view.

2. **Time Series View:** The time series view displays the historical and predicted traffic data for selected road segments. It allows users to analyze the temporal patterns, detect anomalies, and compare different time periods. Users can interactively select specific time intervals, adjust the granularity of the time series, and zoom in to explore detailed traffic variations.

3. **Attention Matrix View:** The attention matrix view visualizes the attention matrices generated by the ST-GRAT model. It represents the weights assigned to different roads and time periods for the predictions. Users can explore the attention matrices to understand how the model assigns importance to different features and identify the roads that significantly influence the predictions. The attention matrix view enhances the interpretability of the ST-GRAT model by providing insights into its internal mechanisms.

By providing these different views and visualizations, AttnAnalyzer enables analysts to interactively explore and interpret traffic data, identify significant features, and gain a deeper understanding of the spatio-temporal dependencies in traffic flow. This supports informed decision-making and facilitates the improvement of traffic management strategies.

4 Model Analysis and Performance Experiment Results

4.1 Case Studies Demonstrating the Effectiveness of AttnAnalyzer

I conduct two case studies on different real-world road networks, namely the urban road network and the highway road network to showcase the effectiveness of AttnAnalyzer in exploring model behaviors and improving performance. These case studies involve:

1. **Comparative Analysis:** I compare the model's behavior of in different traffic road networks, urban and highway, integrated into AttnAnalyzer. I evaluate their predictive accuracy using quantitative metrics and analyze the attention matrices to understand the patterns in capturing spatio-temporal dependencies.

2. **Influential Factors Analysis:** I use AttnAnalyzer to analyze the attention matrices and identify the influential factors in traffic predictions. By examining the weights assigned to different roads and time periods, I can identify the roads that significantly impact traffic conditions and uncover underlying patterns or correlations that contribute to traffic variations. This analysis helps in understanding the important factors that should be considered in traffic management and forecasting.

3. **Model Improvement:** Based on the insights gained from the attention matrices and influential factors analysis, I demonstrate how AttnAnalyzer can be used to improve the performance of traffic forecasting models. By incorporating the identified influential factors or adjusting the attention mechanism, I replace the weight matrix of the models and evaluate their enhanced performance using quantitative metrics. This showcases the potential of AttnAnalyzer in optimizing traffic forecasting models.

These case studies were chosen as representative use-cases based on feedback from domain experts. The first case study focused on the urban road network in Ulsan, South Korea, using DSRC data, while the second case study analyzed the highway road network using METR-LA data. I divided the roads into high error and low error groups based on their Mean Absolute Error (MAE), where the high error group consisted of roads with MAE higher than the third quartile (Q3), and the low error group included roads with MAE lower than the first quartile (Q1).

4.1.1 Data Description

In this work, I use the traffic data of two different road networks–the urban and highway road networks to explore the model's inference process for speed prediction. For the urban road network, I use dedicated short range communication (DSRC) data [24] generated from Ulsan, South Korea (range: 9/1/2017∼12/28/2017), where more than 1.1 million people live with more than 540,000 registered vehicles as of 2017. A total of 116 DSRC sensors are used for data collection, which are installed every 5.7km and cover a total of 68 main roads. For the highway road network, I use the METR-LA data [25], which were collected from 207 loop detectors (range: 3/1/2012∼6/27/2012)

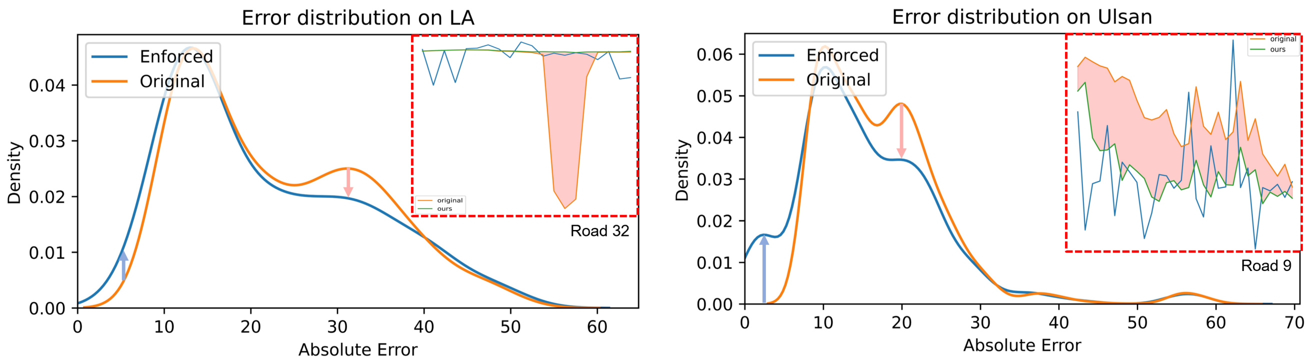


Figure 5: A result of the attention enforcement presents error distribution.

on the highways of Los Angeles. Note that the highway network data I use are the standard benchmark data for traffic forecasting tasks [26, 27, 28]. After discussing with domain experts and reviewing training results, I replace the missing data and explicit errors with historical data. I also use 5-minute aggregated data to mitigate possible effects of outliers, as performed in many previous studies (e.g., [29]).

4.2 Experiment to Improve ST-GRAT using Enforced Weights

The summary of findings using **AttnAnalyzer** are followings:

1. **Finding 1**: When a road's speed experiences significant fluctuations, the model's error rate increases.

2. **Finding 2**: The model frequently struggles to identify relevant references with preceding speed trends.

3. **Finding 3**: The model loses important past self-reference information by dispersing attention to less relevant roads.

To improve the model's performance based on these findings, the researcher employed the attention enforcement method provided in the attention view. Initially, she selected four clusters exhibiting a diagonal pattern in Head 4, as indicated by the green dotted box in Figure 6 (Top). Upon clicking the "Test Alternatives" button, a new view displayed two line charts for performance comparison (Figure 4).

Upon analyzing the chart, notable changes were observed in the original model's performance, indicating improvement. Figure 5 shows how my approach fix the error in general. For instance, when examining the chart of the METR-LA data, there was a decrease in the number of roads with an absolute error (AE) of approximately 30, while the number of roads with an AE of about 5 increased. Similar trends were observed in the Ulsan data, where the number of roads with an AE of around 20 decreased, and the number of roads with an AE of about 5 increased.

These improvements were attributed to the attention enforcement method, which directed the model's focus towards self-reference and roads with preceding speed patterns within the same clusters. By incorporating Findings 2 and 3, the model's performance was significantly enhanced as shown in Table 2.

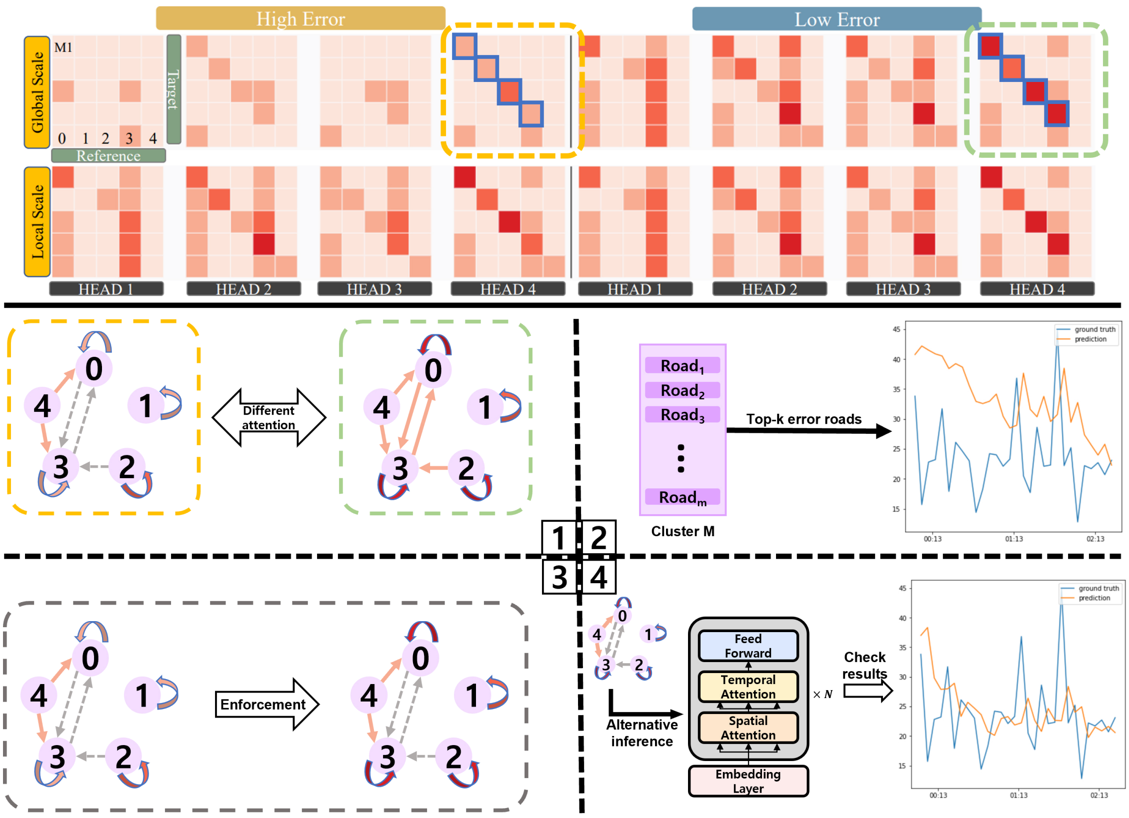


Figure 6: (Top) The head-cluster view with four attention heads, (Bottom) The enforcement process–1) comparing attention between low (left) and high (right) error cases, 2) selecting k highest error roads in each cluster, 3) replacing the attention of the selected roads with that from low error roads, and 4) testing alternatives.

4.3 Domain Expert Feedback and Validation

I actively seek feedback and validation from domain experts to ensure the practical relevance and usability of AttnAnalyzer. I collaborate with traffic machine learning engineers, and data analysts who have expertise in traffic management and forecasting. The domain experts are involved in the design, development, and evaluation stages of AttnAnalyzer.

I conduct user studies with the domain experts to gather their feedback on the usability, effectiveness, and interpretability of AttnAnalyzer's visual analytics features. Through interviews, surveys, and hands-on demonstrations, I collect qualitative feedback on the usefulness of AttnAnalyzer in their workflow, the insights gained from the analysis, and any limitations or areas for improvement. The feedback is invaluable in refining and enhancing the features of AttnAnalyzer to better meet the needs of the target user group.

Furthermore, I validate the outputs and recommendations generated by AttnAnalyzer with domain experts. This validation involves comparing the analysis results and recommendations from AttnAnalyzer with the expertise and domain knowledge of the experts. Their validation helps ensure the accuracy and reliability of AttnAnalyzer in providing insights and actionable information for traffic management and decision-making processes.

The domain expert feedback and validation process strengthen the credibility and effectiveness of AttnAnalyzer as a practical tool for traffic analysis and forecasting.

5 Conclusion

In this dissertation, I have embarked on a comprehensive exploration of improving spatio-temporal data analysis and traffic forecasting through the development and refinement of novel models and visualization tools. My research journey has led me to the creation of the **Spatio-Temporal Graph Attention Network (ST-GRAT)**, a model designed to capture intricate dependencies within spatio-temporal data. With a multi-layered architecture integrating both spatial and temporal attention mechanisms, ST-GRAT demonstrates remarkable capability in discerning patterns and relationships within complex spatio-temporal datasets.

The unveiling of **AttnAnalyzer**, an innovative visualization analytics system, marks a significant stride in the field of model analysis and enhancement. Through AttnAnalyzer, I have empowered researchers and practitioners to gain deeper insights into the inner workings of ST-GRAT. By visualizing attention distributions and exploring interactive representations of the learned dependencies, users can not only understand the model's decisions but also identify areas for improvement and debugging.

With AttnAnalyzer as my guide, I have addressed crucial challenges in ST-GRAT and ushered in a new era of model development. As shown in Table 1, The successful integration of insights gained from AttnAnalyzer has culminated in the creation of **Pattern-Matching Memory Networks (PM-MemNet)** [30]. This advanced iteration of ST-GRAT leverages the power of pattern-based memory representation, enhancing the model's forecasting accuracy and robustness. PM-MemNet stands as a testament to the potential of combining state-of-the-art neural architectures with innovative visualization techniques.

My contributions extend beyond the theoretical realm to practical advancements in the domain of spatio-temporal data analysis. By bridging the gap between model design, analysis, and enhancement, I offer researchers and practitioners a holistic framework for tackling the intricacies of spatio-temporal data. The methodologies and insights presented in this dissertation have the potential to revolutionize various applications, from urban planning to traffic management, by providing accurate predictions and actionable insights.

As I conclude this dissertation, I acknowledge the ever-evolving nature of data analytics and the endless possibilities that lie ahead. The journey presented here is but a stepping stone in the broader landscape of research and innovation. With the foundation laid by ST-GRAT, AttnAnalyzer, and PM-MemNet, I hope to inspire future researchers to explore new horizons, push the boundaries of knowledge, and continue shaping the future of spatio-temporal data analysis and beyond.

References

[1] Cheonbok Park, Chunggi Lee, Hyojin Bahng, Yunwon Tae, Seungmin Jin, Kihwan Kim, Sungahn Ko, and Jaegul Choo. St-grat: A novel spatio-temporal graph attention networks for accurately forecasting dynamically changing road speed. In *Proceedings of the 29th ACM international conference on information & knowledge management*, pages 1215–1224, 2020.

[2] Ameet Talwalkar, Sanjiv Kumar, and Henry Rowley. Large-scale manifold learning. In *2008 IEEE Conference on Computer Vision and Pattern Recognition*, pages 1–8. IEEE, 2008.

[3] Seungmin Jin, Hyunwook Lee, Cheonbok Park, Hyeshin Chu, Yunwon Tae, Jaegul Choo, and Sungahn Ko. A visual analytics system for improving attention-based traffic forecasting models. *IEEE transactions on visualization and computer graphics*, 29(1):1102–1112, 2022.

[4] David Alexander Tedjopurnomo, Zhifeng Bao, Baihua Zheng, Farhana Murtaza Choudhury, and Alex Kai Qin. A survey on modern deep neural network for traffic prediction: Trends, methods and challenges. *IEEE Transactions on Knowledge and Data Engineering*, 34(4):1544–1561, 2020.

[5] Renhe Jiang, Du Yin, Zhaonan Wang, Yizhuo Wang, Jiewen Deng, Hangchen Liu, Zekun Cai, Jinliang Deng, Xuan Song, and Ryosuke Shibasaki. Dl-traff: Survey and benchmark of deep learning models for urban traffic prediction. In *Proceedings of the 30th ACM international conference on information & knowledge management*, pages 4515–4525, 2021.

[6] Yaguang Li and Cyrus Shahabi. A brief overview of machine learning methods for short-term traffic forecasting and future directions. *Sigspatial Special*, 10(1):3–9, 2018.

[7] Chunggi Lee, Yeonjun Kim, Seungmin Jin, Dongmin Kim, Ross Maciejewski, David Ebert, and Sungahn Ko. A visual analytics system for exploring, monitoring, and forecasting road traffic congestion. *IEEE transactions on visualization and computer graphics*, 26(11):3133–3146, 2019.

[8] Tianyi Zhang, Cuiyun Gao, Lei Ma, Michael Lyu, and Miryung Kim. An empirical study of common challenges in developing deep learning applications. In *2019 IEEE 30th International Symposium on Software Reliability Engineering (ISSRE)*, pages 104–115. IEEE, 2019.

[9] Mikhail Belkin, Siyuan Ma, and Soumik Mandal. To understand deep learning we need to understand kernel learning. In *International Conference on Machine Learning*, pages 541–549. PMLR, 2018.

[10] Chuanpan Zheng, Xiaoliang Fan, Cheng Wang, and Jianzhong Qi. Gman: A graph multi-attention network for traffic prediction. In *Proceedings of the AAAI conference on artificial intelligence*, volume 34, pages 1234–1241, 2020.

[11] Rongzhou Huang, Chuyin Huang, Yubao Liu, Genan Dai, and Weiyang Kong. Lsgcn: Long short-term traffic prediction with graph convolutional networks. In *IJCAI*, volume 7, pages 2355–2361, 2020.

[12] Vincent Le Guen and Nicolas Thome. Shape and time distortion loss for training deep time series forecasting models. *Advances in neural information processing systems*, 32, 2019.

[13] Lei Zhang, Kaiqun Fu, Taoran Ji, and Chang-Tien Lu. Granger causal inference for interpretable traffic prediction. In *2022 IEEE 25th International Conference on Intelligent Transportation Systems (ITSC)*, pages 1645–1651. IEEE, 2022.

[14] James Thomas and Kristin Cook. *Illuminating the Path: The Research and Development Agenda for Visual Analytics*. National Visualization and Analytics Ctr, 2005.

[15] Zhuoning Yuan, Xun Zhou, and Tianbao Yang. Hetero-convlstm: A deep learning approach to traffic accident prediction on heterogeneous spatio-temporal data. In *Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*, pages 984–992, 2018.

[16] Gianni Brauwers and Flavius Frasincar. A general survey on attention mechanisms in deep learning. *IEEE Transactions on Knowledge and Data Engineering*, 2021.

[17] Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N Gomez, Łukasz Kaiser, and Illia Polosukhin. Attention is all you need. *Advances in neural information processing systems*, 30, 2017.

[18] Weiwei Jiang and Jiayun Luo. Graph neural network for traffic forecasting: A survey. *Expert Systems with Applications*, page 117921, 2022.

[19] Cen Chen, Kenli Li, Sin G Teo, Xiaofeng Zou, Kang Wang, Jie Wang, and Zeng Zeng. Gated residual recurrent graph neural networks for traffic prediction. In *Proceedings of the AAAI conference on artificial intelligence*, volume 33, pages 485–492, 2019.

[20] Qin Zhang, Keping Yu, Zhiwei Guo, Sahil Garg, Joel JPC Rodrigues, Mohammad Mehedi Hassan, and Mohsen Guizani. Graph neural network-driven traffic forecasting for the connected internet of vehicles. *IEEE Transactions on Network Science and Engineering*, 9(5):3015–3027, 2021.

[21] Gennady Andrienko, Natalia Andrienko, Wei Chen, Ross Maciejewski, and Ye Zhao. Visual analytics of mobility and transportation: State of the art and further research directions. *IEEE Transactions on Intelligent Transportation Systems*, 18(8):2232–2249, 2017.

[22] Chenyi Chen, Yin Wang, Li Li, Jianming Hu, and Zuo Zhang. The retrieval of intra-day trend and its influence on traffic prediction. *Transportation research part C: emerging technologies*, 22:103–118, 2012.

[23] Shengmin Guo, Dong Zhou, Jingfang Fan, Qingfeng Tong, Tongyu Zhu, Weifeng Lv, Daqing Li, and Shlomo Havlin. Identifying the most influential roads based on traffic correlation networks. *EPJ Data Science*, 8(1):1–17, 2019.

[24] John B Kenney. Dedicated short-range communications (dsrc) standards in the united states. *Proceedings of the IEEE*, 99(7):1162–1182, 2011.

[25] Zonghan Wu, Shirui Pan, Guodong Long, Jing Jiang, and Chengqi Zhang. Graph wavenet for deep spatial-temporal graph modeling. *arXiv preprint arXiv:1906.00121*, 2019.

[26] Yaguang Li, Rose Yu, Cyrus Shahabi, and Yan Liu. Diffusion convolutional recurrent neural network: Data-driven traffic forecasting. *arXiv preprint arXiv:1707.01926*, 2017.

[27] Bing Yu, Haoteng Yin, and Zhanxing Zhu. Spatio-temporal graph convolutional networks: A deep learning framework for traffic forecasting. In *Proceedings of the 27th International Joint Conference on Artificial Intelligence (IJCAI)*, 2018.

[28] Tanwi Mallick, Prasanna Balaprakash, Eric Rask, and Jane Macfarlane. Transfer learning with graph neural networks for short-term highway traffic forecasting. In *2020 25th International Conference on Pattern Recognition (ICPR)*, pages 10367–10374. IEEE, 2021.

[29] Jiani Zhang, Xingjian Shi, Junyuan Xie, Hao Ma, Irwin King, and Dit-Yan Yeung. Gaan: Gated attention networks for learning on large and spatiotemporal graphs. *arXiv preprint arXiv:1803.07294*, 2018.

[30] Hyunwook Lee, Seungmin Jin, Hyeshin Chu, Hongkyu Lim, and Sungahn Ko. Learning to remember patterns: Pattern matching memory networks for traffic forecasting. In *International Conference on Learning Representations*, 2021.