Exploring Complex Graph Representations of Highway Networks for Traffic Speed Forecasting

Ricky Miura
rmiura@gmail.com

Gita Anand
gianand@ucsd.edu

Sheena Patel
s4patel@ucsd.edu

Mentor: Gal Mishne
gmishne@ucsd.edu

Mentor: Yusu Wang
yuw122@ucsd.edu

UC San Diego
HALICIOĞLU DATA SCIENCE INSTITUTE

Motivation

- Traffic is an enduring issue given the extensive automobile use today but can be improved through modeling traffic flow prediction
- Challenges of predicting future traffic conditions include difficulty of factoring spatial and temporal dependencies in traffic dynamics

Our proposed model, a Spatial-Temporal Graph Attention Network (ST-GAT) aims to explore the following:

- Identify the time frame that yields the most accurate predictions of traffic speeds (15 min/30 min/45 min)
- Explore various node feature and edge type combinations in order to determine what factors are most influential on predictive performance

Contribution

- **Data Collection and Preprocessing**: Manually extract data off Caltrans PeMS and convert tabular data into graph form
- **Model Implementation and Evaluation**: Create 12 distinct models with varying node features, and either learned or unlearned edge types of varying quantities and compare performance using MAPE, MAE, RMSE metrics

Data Collection

- Leveraged **California's Performance Measurement System** (PeMS), to manually extract data from sensors stations on San Diego County's 3 main highway systems over a restricted time period of 14 days with 5-minute intervals
- Found connectivity and calculated distances between sensors using longitude and latitude coordinates in order to construct graphs

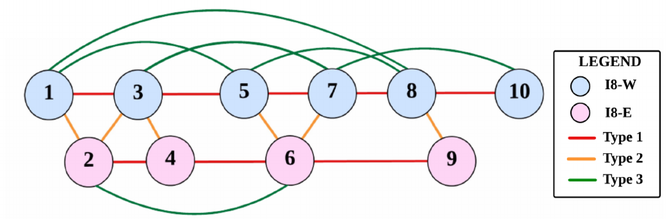
Graph	Node Features	Edge Types Included
1	Speed	1
2	Speed	1, 2
3	Speed	1, 2, 3
4	Speed, Lanes, Day of Week, Hour of Day	1
5	Speed, Lanes, Day of Week, Hour of Day	1, 2
6	Speed, Lanes, Day of Week, Hour of Day	1, 2, 3

Graph Construction

Node Features

- **Speed Window**: Captures dynamic speed variations by utilizing a window of previous aggregated speeds from specific time intervals
- **Lane Count**: Identifies the constant number of lanes adjacent to each sensor as a critical factor for understanding traffic flow dynamics and bottleneck phenomena
- **Day of Week**: Illustrates significant fluctuations in traffic patterns across different days, incorporating one-hot encoding for each day of the week
- **Hour of Day**: Incorporates the time of day, using a sinusoidal and cosine transformation to capture the cyclical nature of traffic patterns, notably during rush hours

Edge Types

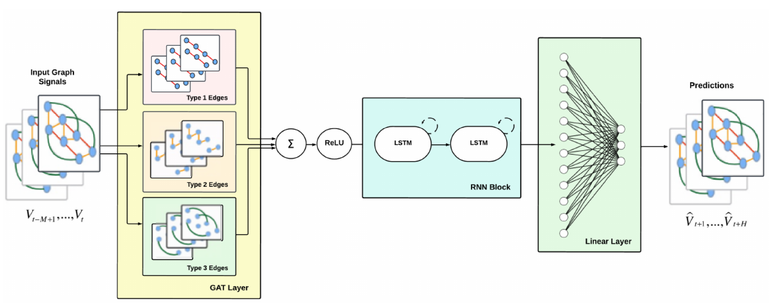


- **Type 1 Edge**: Establishes linear connections between sensors, ensuring sequential linkage and proximity-based connectivity within the network, represented by matrix W1
- **Type 2 Edge**: Accounts for external influences on a sensor's highway, capturing minor traffic patterns from alternate routes, represented by matrix W2
- **Type 3 Edge**: Captures long-range traffic dynamics within a sensor's highway, forming skip connections to mitigate oversmoothing, represented by matrix W3

Model Construction

Spatial-Temporal Graph Attention Network (ST-GAT)

- **GAT Layer**: Leverages attention mechanisms to weigh the influence of neighboring nodes differently, capturing the spatial dependencies within the graph structure
- **LSTM Layer**: Uses Long Short-Term Memory units that are adept at capturing temporal dynamics, allowing the model to remember long-term dependencies and discern sequential patterns over time



- **ST-GAT Single Edge**: Processes the graph by considering all edge types as equivalent, without distinguishing between them, which simplifies the model while still allowing for the integration of various combinations of edge types to predict traffic speeds
- **ST-GAT Edge Type**: Enhances the model's ability to understand the traffic network's spatial-temporal dynamics by differentiating between edge types and assigning unique weights to each, enabling a more detailed and nuanced traffic speed prediction

Results

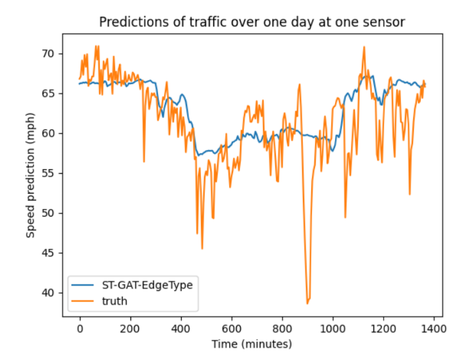
Evaluations for ST-GAT Single Edge

Graph	15 min			30 min			45 min		
	RMSE	MAE	MAPE (%)	RMSE	MAE	MAPE (%)	RMSE	MAE	MAPE (%)
1	3.77	2.01	4.67	3.90	2.06	4.83	4.04	2.13	5.03
2	3.89	2.04	4.73	3.97	2.08	4.90	4.07	2.11	5.03
3	3.85	2.01	4.67	3.94	2.05	4.82	4.08	2.15	5.08
4	3.79	2.05	4.68	3.91	2.08	4.87	4.07	2.16	5.06
5	3.91	2.07	4.79	4.00	2.14	5.02	4.16	2.13	5.00
6	3.89	2.03	4.72	4.07	2.09	4.93	4.03	2.06	4.93

Evaluations for ST-GAT Edge Type

Graph	15 min			30 min			45 min		
	RMSE	MAE	MAPE (%)	RMSE	MAE	MAPE (%)	RMSE	MAE	MAPE (%)
1	3.79	1.98	4.62	3.97	2.03	4.87	3.98	2.05	4.91
2	3.78	2.01	4.69	3.94	2.04	4.89	4.06	2.13	5.06
3	3.87	2.03	4.66	4.11	2.13	5.02	4.11	2.13	5.02
4	3.82	2.02	4.71	3.94	2.03	4.86	4.03	2.12	4.99
5	3.93	2.02	4.75	3.97	2.07	4.86	4.04	2.12	5.01
6	3.85	2.06	4.71	4.01	2.09	4.87	4.10	2.10	5.00

- Forecasting traffic speeds for the upcoming 15 minutes consistently outperformed predictions for longer timeframes, demonstrating the model's strength in near-term accuracy
- The inclusion of extra features led to a dip in performance, likely attributed to the over-generalization of time to hourly increments rather than to more precise intervals
- Graphs using ST-GAT Edge Type generally performed better than its ST-GAT Single Edge counterpart which shows the importance of differentiating edge types
- Graphs containing Type 1 edges performed the best which may be due to the close proximity of our sensors and nearest neighbor speeds providing the most reliable information for predicting a target sensor's speed



Future Work

- Make data collection process automated for more efficient retrieval of sensor data
- Collect sensor data that extends a longer time frame to increase sample size for model training
- Integrate other highways in San Diego County to capture more highway relations
- Explore advanced optimization techniques such as Grid Search Cross Validation to optimize hyperparameters
- Incorporate additional features that factor into traffic patterns such as weather and lane closures to enrich feature vectors
- Integrate real-time feedback mechanisms into prediction models to enable continuous learning and adaptation to changing traffic and feature conditions