

Drive Me Crazy

Technical details	
Submit files	dataset_analysis.ipynb - presentation.txt - lookerstudio_url.txt - drive_me_crazy_tradi.ipynb - drive_me_crazy_pdformer.ipynb
Languages	It needs to be completed in the language you are working on right now. If you are doing Bootcamp Javascript, then javascript (file extension will be .js). If you are doing Bootcamp Ruby, then Ruby (file extension will be .rb). It goes the same for Python, Java, C++, Rust, ...

The objective of this project is to develop a robust traffic flow prediction system using the "Propagation Delay-Aware Dynamic Long-Range Transformer" method. The system will utilize data from PeMS04, PeMS07, PeMS08, and NYCTaxi datasets to forecast traffic flow patterns accurately in urban areas while considering propagation delays. Additionally, one of the deliverables of the project will be the development of an interactive dashboard using Looker Studio to visualize and present the traffic flow predictions.

Objectives

The main objectives of the project are as follows:

Download the PeMS04, PeMS07, PeMS08, and NYCTaxi datasets. You will find them in the Internet. Implement a "traditional" prediction model on those datasets to predict traffic.

Implement the "Propagation Delay-Aware Dynamic Long-Range Transformer" and you will test it on those models. (<https://arxiv.org/pdf/2301.07945v2.pdf>) Evaluate the model's predictive performance and compare it against traditional traffic prediction methods. Investigate the impact of propagation delays on the accuracy of traffic flow predictions using real-world data.

Analyze the efficiency of dynamic long-range transformers in capturing temporal and spatial dependencies across diverse urban traffic scenarios. Develop an interactive dashboard using Looker Studio to visualize and present the traffic flow predictions.

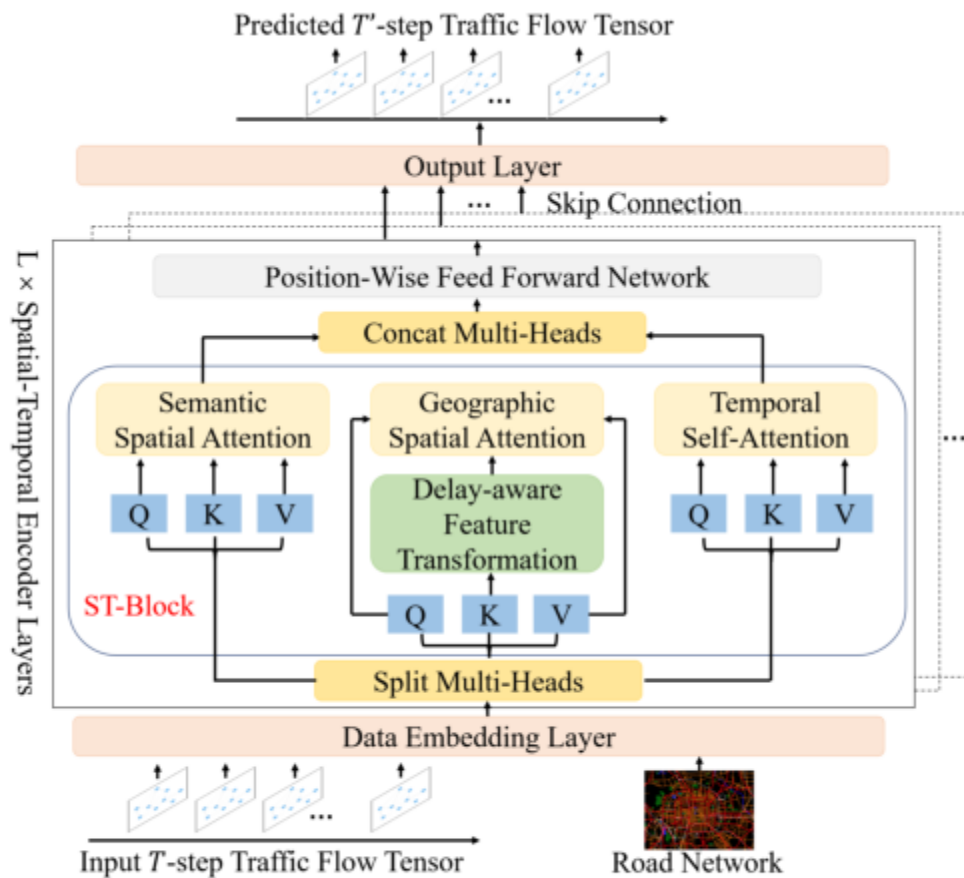


Figure 2: The Overall Framework of PDFormer.

Methodology

The project will follow these essential steps:

- **Data Integration:** Combine the PeMS04, PeMS07, PeMS08, and NYCTaxi datasets to create a unified traffic data collection.
- **Data Preprocessing:** Cleanse, aggregate, and preprocess the integrated data to remove inconsistencies and prepare it for training.
- **Model Adaptation:** Customize the "Propagation Delay-Aware Dynamic Long-Range Transformer" model to accommodate the characteristics of the combined dataset.
- **Model Training:** Train the adapted model using appropriate data science libraries, considering relevant traffic features and historical patterns.

- **Model Evaluation:** Assess the model's predictive accuracy using metrics such as Mean Absolute Error (MAE) and Root Mean Squared Error (RMSE).
- **Comparison Analysis:** Compare the performance of the proposed model with traditional traffic prediction approaches using the same datasets.
- **Propagation Delay Investigation:** Analyze the influence of propagation delays on traffic flow predictions and identify potential challenges or bias.
- **Transformer Efficiency Analysis:** Investigate the effectiveness of dynamic long-range transformers in capturing traffic patterns across different datasets.
- **Looker Studio Dashboard:** Design and build an interactive Looker Studio dashboard to visualize and present the traffic flow predictions and model evaluation results.

Deliverables

We would like you to complete the following:

- An analysis of the dataset.
- A traditional traffic flow prediction model. (A jupyter notebook.)
- An implemented "Propagation Delay-Aware Dynamic Long-Range Transformer" model tailored for accurate traffic flow prediction. (A jupyter notebook.)
- Analysis of the impact of propagation delays on traffic flow prediction accuracy. (A paper about your findings)
- An interactive Looker Studio dashboard showcasing traffic flow predictions and model evaluation results.
- A presentation reports comparing the model's performance against traditional traffic prediction methods using various datasets.
- Insights into the efficiency of dynamic long-range transformers in capturing diverse urban traffic patterns.
- Make sure you don't push the dataset(s) to git, it will be an automatic rejection.