Advanced Traffic Volume Estimation Using Machine Learning

Advanced Traffic Volume Estimation Using Machine Learning

Abstract:

Urban traffic congestion is a growing challenge that demands innovative solutions for effective traffic management and planning. This project explores the application of machine learning and deep learning techniques for advanced traffic volume estimation. Using real-time and historical traffic data, we implement and compare supervised learning models (Random Forest, Gradient Boosting, and SVR) and deep learning models (LSTM networks) to identify the most effective approach for accurate traffic prediction.

1. Introduction:

With rapid urbanization, cities are experiencing a surge in vehicular movement, leading to congestion, pollution, and increased travel times. Traditional traffic volume estimation methods often fall short in dynamic environments. Machine learning offers robust and adaptive solutions by learning from vast datasets, making it suitable for modern intelligent transportation systems (ITS).

2. Literature Review:

Previous studies have utilized linear regression and time-series models like ARIMA, but these methods often fail to capture non-linear and temporal dependencies. Supervised machine learning models have demonstrated improved accuracy, while deep learning, particularly LSTM networks, shows promise in capturing complex temporal patterns.

3. Data Collection and Preprocessing:

We utilized publicly available datasets containing:

- Traffic sensor data

- Weather conditions (temperature, precipitation)
- Timestamps (hour, day, week)
- Holiday and weekend indicators

Preprocessing involved handling missing values, normalization, and resampling data into 15-minute intervals. Feature engineering was performed to improve model input quality.

4. Proposed Methodology:

4.1 Supervised Machine Learning Models:

- Random Forest Regressor: Aggregates predictions from multiple decision trees for robust estimation.
- Gradient Boosting: Optimizes prediction through iterative improvement and error minimization.
- Support Vector Regression (SVR): Uses kernel functions for high-dimensional regression tasks.

4.2 Deep Learning Model:

- LSTM Networks: Captures long-term dependencies in time-series data using memory cells and gates.
- Model architecture includes input layers, one or more LSTM layers, and a dense output layer. Hyperparameters like sequence length, neuron count, and dropout rate were optimized through grid search.

5. Experimental Results:

Models were evaluated using Mean Absolute Error (MAE) and Root Mean Square Error (RMSE):

- Random Forest achieved the lowest error among supervised models.

- LSTM outperformed all other models, showing strong alignment with actual traffic volumes, especially during peak times.

6. Conclusion and Future Work:

The combination of machine learning and deep learning techniques significantly enhances traffic volume prediction. LSTM networks show the best performance, making them suitable for real-time applications. Future work will focus on deploying these models in live environments and exploring hybrid architectures (e.g., CNN-LSTM) for spatial-temporal analysis.

Keywords: Traffic Volume, Machine Learning, Deep Learning, Random Forest, LSTM, Traffic Prediction, Time Series, Intelligent Transportation Systems