Predicting Near-Road Air Quality Using Artificial Neural Networks: Exploring Traffic-Related Influences

YIHONG NING

PH.D. STUDENT IN CIVIL ENGINEERING

12/11/2024

Motivation and Background

Air Quality and Health Risks:

- Poor air quality, particularly near roadways, is a significant health concern.
- Pollutants like NO_2 and PMs linked to respiratory and cardiovascular diseases.
- Traffic emissions are primary contributors.

Challenge with Traditional Models:

- Traditional models (e.g., linear regression) often fail to capture the complex, non-linear interactions between traffic behaviors and environmental factors affecting air quality.

Objective:

- Develop an Artificial Neural Network (ANN) model that can predict near-road air quality more accurately by capturing complex relationships in data that traditional models might miss.

Data Collection and Preparation

Location: Data collected from Taylor Street, Columbia, SC.

Variables:

- Response Variables: PM1.0, PM2.5, PM10, NO_2.
- Predictors: Traffic (counts of cars, trucks, speed, gaps) and Environmental (temperature, humidity, pressure).

Data Preparation:

- Data cleaning and integration at 15-minute intervals.
- Log transformation applied to pollutant variables to reduce skewness.

ANN Model

Structure:

- **Input Layer:** Includes predictors after preparation.
- **Hidden Layers:** 3 layers with 128, 64, and 32 neurons, ReLU activation, and dropout regularization. Dropout rate of 0.3 after the first two hidden layers.
- **Output Layer:** Single neuron for regression to predict pollutant concentrations.

Training Strategy:

- 80% data for training, 20% data for testing.
- Cross-validation with early stopping to prevent overfitting.

Evaluation Metrics:

- R^2 ,RMSE, LIME, Cross-validation.
- Compared with Multiple Linear Regression (MLR) and Bayesian Model Averaging (BMA) to validate performance.

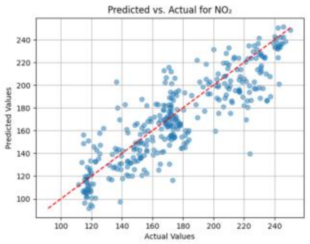
ANN Performance

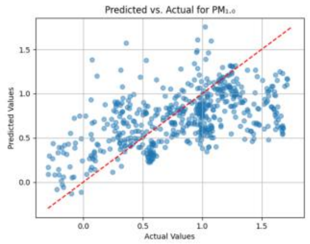
Models	ANN		MLR		BMA	
Air Pollutant	R^2	**RMSE**	R^2	**RMSE**	R^2	**RMSE**
PM1.0	0.881	0.164	0.320	8.410	0.443	0.364
PM2.5	0.889	0.157	0.321	11.366	0.464	0.356
PM10	0.884	0.159	0.321	11.951	0.473	0.352
NO_2	0.868	15.087	0.832	15.836	0.833	14.812

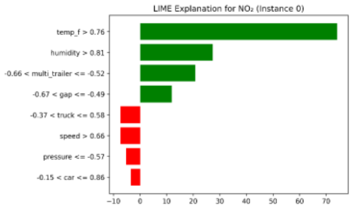
Findings:

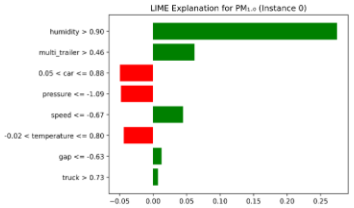
- ANN's superior performance suggests that nonlinear, data-driven methods can more effectively learn intricate patterns in near-road pollutant dynamics.
- BMA's uncertainty quantification remains valuable.

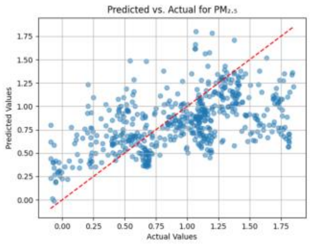
ANN Performance

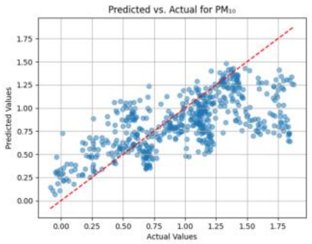


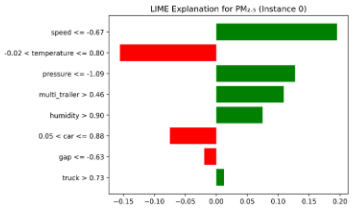


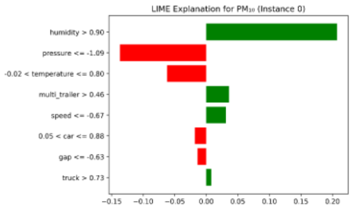












Findings:

- The predicted vs. actual plots show that overall performance remains robust.
- LIME plots indicate:
  - Heavy duty trucks are significant contributors to NO2 and PMs.
  - Temperature and humidity play substantial roles on PMs.
  - Speed and vehicle density trend to influence NO2.

Limitations and Future work

Limitations:

- Potential overfitting
- High quality data dependency

Future work:

- Integrate ANN with probabilistic approaches like Bayesian model.