REAL-TIME TRAFFIC CONGESTION PREDICTION AND MANAGEMENT

Eric Pihlstrom
CSPB
University of Colorado
Boulder CO, USA

Problem Statement/Motivation

Urban centers globally face the perennial challenge of traffic congestion, with its associated economic, environmental, and societal costs. The drive towards sustainable urban mobility necessitates innovative approaches to predict and manage traffic flow dynamically. Traditional traffic management systems, while effective to a degree, often fall short in adapting to real-time conditions and predicting congestion with high accuracy. This project is motivated by the potential of advanced data analytics and machine learning to revolutionize traffic congestion prediction and management. By leveraging datasets, the proposed model aims to provide actionable insights for traffic planners and policymakers, thereby enhancing urban mobility and reducing traffic-related environmental impacts.

Literature Survey

Real-time Traffic Congestion Prediction Using Big Data and Machine Learning Techniques: This study, conducted by Chawla et al., proposes an intelligent real-time traffic model designed to assist urban populations by predicting traffic information and the likelihood of road accidents, ultimately aiming to reduce carbon dioxide emissions and improve transportation quality. The model integrates data from road sensors and various sources, processed through streaming analytics platforms utilizing deep learning and machine learning techniques. The study highlights the potential of IoT sensor data alongside other data sources in enhancing traffic management and strategic decision-making for transit agencies and public safety departments. The research acknowledges the impact of the COVID-19 pandemic on traffic patterns, suggesting the model's predictions for post-February 2020 might be less accurate due to the pandemic's effects.

Proposed Work

Data Collection:

- Utilize datasets from:
    - Kaggle: Traffic Prediction Dataset.
    - Data.gov: Highway Performance Monitoring System (HPMS), Traffic Flow Management Data Integrated Terminal Weather.

Data Preprocessing:

- Clean data to correct errors, remove outliers, and fill missing values.
- Aggregate and normalize data formats from different sources.
- Create time-series datasets for analysis.

Data Integration:

- Combine datasets based on time and location for a comprehensive view of traffic congestion factors.

Feature Engineering:

- Develop features that reflect traffic flow dynamics, weather conditions, and event impacts.

Methodology Difference:

- This project differs from previous work by integrating a wider range of data sources and employing more sophisticated data preprocessing and integration methods.

Dataset Details

- Kaggle Traffic Prediction Dataset URL: https://www.kaggle.com/datasets/fedesoriano/trafficprediction-dataset
- Data.gov Datasets URLs: HPMS and Traffic Flow Management Data Integrated Terminal Weather.

Evaluation Methods

1. The model's performance will be evaluated through benchmarking against existing traffic prediction models and assessing its adaptability to new data. Metrics such as accuracy, precision, recall, and F1 score will be used to evaluate prediction performance. Additionally, the model's ability to integrate and adapt to new data streams will be a key focus of the evaluation, ensuring its long-term applicability in dynamic urban environments.

Tools

For the "Real-Time Traffic Congestion Prediction and Management" project, several tools and software libraries will be pivotal in data processing, analysis, visualization, and model development:

- **Python**: A versatile programming language that serves as the backbone for the majority of data analysis and machine learning tasks. Python's extensive ecosystem includes numerous libraries that will be essential for this project.
  - **Pandas**: For data manipulation and analysis. Pandas provide data structures and operations for manipulating numerical tables and time series.
  - **NumPy**: Fundamental package for scientific computing with Python. It provides support for large, multi-dimensional arrays and matrices, along with a collection of mathematical functions to operate on these arrays.
  - **Scikit-learn**: An open-source machine learning library for Python. It features various classification, regression, and clustering algorithms including support vector machines, random forests, gradient boosting, k-means, and DBSCAN, and is designed to interoperate with NumPy and Pandas.
  - **Matplotlib** and **Plotly**: For data visualization. These libraries allow for the creation of static, animated, and interactive visualizations in Python.
- **Jupyter Notebooks**: An open-source web application that allows you to create and share documents that contain live code, equations, visualizations, and narrative text. Jupyter Notebooks will be used for interactive analysis and reporting.
- **PostgreSQL** or similar database systems (for data storage and spatial queries): To store, query, and manage the traffic data, especially if dealing with geospatial data, PostGIS adds support for geographic objects to the PostgreSQL database, turning it into a spatial database.

Milestones

The project will be structured around several key milestones to ensure timely progress and successful completion. Each milestone represents a significant phase in the project's development cycle:

1. **Project Kick-off and Preliminary Research (March 4)**: Finalize project scope, objectives, and perform an extensive literature review to refine the methodology.

2. **Data Collection and Initial Processing (March 18)**: Acquire traffic, weather, and other relevant datasets. Begin preprocessing to clean and standardize the data formats.

3. **Data Integration and Feature Engineering (April 15)**: Integrate data from different sources and develop new features that capture the dynamics influencing traffic congestion.

4. **Model Development and Training (April 22)**: Experiment with different algorithms to find the most effective approach.

5. **Model Evaluation and Refinement (April 22)**: Benchmark the model's performance against existing solutions and refine based on feedback and results.

6. **Deployment of Prediction Model (May 2): Finalize project and turn in.**

Resources:

Chawla, P., Hasurkar, R., Bogadi, C.R., Korlapati, N.S., Rajendran, R., Ravichandran, S., Tolem, S.C., & Gao, J.Z. (2024). Real-time traffic congestion prediction using big data and machine learning techniques. *World Journal of Engineering*, Vol. 21 No. 1, pp. 140-155