Homework 3: Multimodal Machine Learning for Transit Data

Riccardo Fiorista — riccardo-uai@mit.edu

August, 2025

Lecture Summary

Lecture 3 highlighted real-world applications of AI in public transportation, drawing on case studies from MIT's decade-long collaborations with major transit authorities worldwide (TfL, CTA, Hong Kong MTR, MBTA, Singapore LTA, WMATA). The lecture illustrated how AI progresses from theory to practice across three domains: **performance analysis** (origin–destination flows, demand–capacity balance), **behavioral insights** (customer segmentation, crowding effects), and **decision support** (simulation, operations control, crew scheduling).

- **Case Study 1 – Platform Crowding Detection:** Hong Kong MTR's automation of manual passenger counts using AFC/AVL data and AI-based estimation; WMATA's integration of CCTV computer vision with operational data, yielding a 30% accuracy improvement, combined with selective resource allocation via reinforcement learning.

- **Case Study 2 – Bus Dispatch Control:** The Chicago Transit Authority's reinforcement learning system for terminal dispatching, reducing excess waiting time by 37% and crowding by 5–20%. The system explicitly accounted for driver compliance (averaging 40–50%) and real-world operational constraints.

- **Case Study 3 – Sentiment Analysis:** WMATA's conversion of unstructured social media messages into structured operational insights. Tens of thousands of posts were analyzed for mode, route, location, time, sentiment, and demographic attributes, enabling systematic complaint tracking and policy evaluation.

Across all examples, the lecture emphasized human-in-the-loop system design, ethical AI practices (de-identification and privacy protection), and the progression from retrospective analysis to predictive analytics enabled by multimodal data integration.

In this recitation, we will focus on Case Studies 1 and 3, guiding you through minimal viable implementations of their core AI functionalities. The exercises are organized in two sections: first on sentiment analysis (Case Study 3), and then on platform crowding detection (Case Study 1). All code is provided in a Google Colab notebook, enabling you to both read and execute it in a reproducible environment.

Resources

Colab Notebook: Access the interactive implementation at:
`https://colab.research.google.com/github/jtl-transit/UAI-Transportation-2025/blob/main/recitations/recitation_3/recitation_3_code.ipynb`

Exercise 1: Sentiment Analysis with MetRoBERTa

> ## Apply transformer-based NLP models to analyze public opinions on a transit agency and extract operational insights

This exercise implements the sentiment analysis methodology from Case Study 3, where WMATA processes large volumes of social media messages to extract structured operational data. You'll work with three pre-trained BERT models: one trained to detect *sentiment*, one for *irony*, and one which we fine-tuned to identify the underlying, transit-related *topics* of the post. This exercise is designed to help you understand how unstructured narrative data can be transformed into actionable insights.

Technical Implementation

The respective notebook section implements the following pipeline:

(a) **Data Loading:** WMATA Twitter dataset (5,000 messages) with preprocessing

(b) **Model Architecture:** MetRoBERTa - a BERT model fine-tuned on transportation domain text

(c) **Feature Extraction:** Automated classification of:

- Mode (bus, train, metro)
- Route identification
- Location/stop extraction
- Time period inference
- Sentiment classification (positive, negative, sarcastic, neutral)
- Topic categorization (schedule, cleanliness, safety, facilities)

(d) **Analysis Pipeline:** Aggregate insights and trend detection

Instructions

(a) **Execute the Notebook:** Run through all cells in the Colab notebook, observing the model's performance on different types of transit-related text

(b) **Analyze Model Outputs:** Examine how the transformer model handles:

- Sarcasm detection (e.g., "nice trash pile to start my morning")
- Multi-attribute extraction from complex sentences
- Domain-specific transportation terminology

(c) **Performance Assessment:** Evaluate the model's accuracy across different categories and identify potential failure modes

(d) **Data Quality Analysis:** Consider how the quality and representativeness of social media data affects operational decision-making

Goal: Reflect on your observations and analysis of the model's performance, particularly noting how well it captures the nuances Professor Zhao discussed regarding converting this type of data into structured operational insights. We will follow up with some multiple-choice questions throughout the recitation videos.

Exercise 2: Computer Vision for Crowding Detection

> ## Implement semantic segmentation approaches for automated passenger density estimation in transit stations

This exercise replicates the computer vision methodology from Case Study 1, where WMATA integrated CCTV analysis with operational data to improve crowding predictions by 30%. You'll work with station video data that we manually sampled in the city of Vienna, Austria, to understand how image segmentation can estimate passenger density.

Technical Implementation

In the same Colab notebook, you will find a section that contains the computer vision implementation with:

(a) **Video Data:** Vienna Metro station footage showing varying crowd densities over 10 minutes

(b) **Segmentation Model:** Pre-trained semantic segmentation network for person detection

(c) **Density Estimation:** Area-based crowd density calculation (not individual tracking)

(d) **Temporal Analysis:** Frame-by-frame crowding progression over time

Instructions

(a) **Inspect Data:** Observe what data is given to you, how it is loaded and preprocessed

(b) **Run Segmentation Pipeline:** Execute the computer vision cells and observe how the model segments crowd areas

(c) **Analyze Density:** Examine how pixel-based area calculations translate to passenger density estimates

(d) **Continuous Inference:** Understand why we modified our inference pipeline from a per-frame (every image in a video) to a one-image-per-second approach

(e) **Evaluate Computational Trade-offs:** Consider the balance between *accuracy* (i.e., *how often to sample an image* and at what *threshold*) and *computational cost* that Professor Zhao discussed. In this particular case, can you think of a simpler approach?

(f) **Integration Challenges:** Think about how you would combine the continuous insight extracted from the vision data with other operational data sources (train delays, smart card usage, etc.)

Goal: Reflect on the computer vision approach, including its strengths, limitations, and practical deployment considerations for real-world transit operations. We will follow up with some multiple-choice questions throughout the recitation videos.

Exercise 3: Applying Multimodal AI to Your Domain

> ## Design a multimodal AI system for your field/industry
> ## that integrates text and visual data sources

Drawing from Professor Zhao's three case studies and your hands-on experience with the NLP and computer vision implementations, conceive, either mentally or by reusing some of the code, a comprehensive AI system for your chosen domain, potentially the one you worked on in Homework 1, that demonstrates the principles of grounded AI deployment.

System Design Requirements

Your proposed system should address:

Problem Definition:

- Identify a specific operational challenge in your field analogous to crowding detection, service reliability, or sentiment analysis

- Define clear success metrics and operational constraints

- Consider stakeholder needs across operational, tactical, and strategic levels

Multimodal Data Integration:

- **Structured Data:** Equivalent to fare collection (Automatic Fare Collection, AFC) or operational log (such as Automatic Vehicle Location, AVL) systems in your domain

- **Text Data:** Social media, customer feedback, or operational communications

- **Visual Data:** Security cameras, satellite imagery, or sensor networks

- **Contextual Data:** Weather, events, temporal patterns

Technical Architecture:

- **Data Pipeline:** How the different data sources are ingested and processed

- **AI Components:** Which of the five AI functions (represent, predict, explain, control, create) your system employs

- **Human-in-the-Loop Design:** Where human oversight and decision-making remain critical

- **Computational Efficiency:** When to use expensive AI components vs. simpler methods

Implementation Considerations

Address the practical challenges Professor Zhao highlighted:

(a) **Compliance and Adoption:** How will you handle variable user compliance (like the 40–50% driver compliance in Chicago)?

(b) **Ethical Deployment:** What privacy protections and de-identification measures are necessary?

(c) **Organizational Integration:** How would your system fit into existing workflows and institutional culture?

(d) **Scalability:** Path from pilot implementation to system-wide deployment

(e) **Cross-Cultural Learning:** How could your approach be adapted across different contexts or organizations?

Reflection on Grounded AI Principles

Demonstrate how your system embodies the grounded AI characteristics:

- **Problem-driven:** Addresses real operational needs, not just technical capabilities

- **Context-aware:** Considers institutional, cultural, and social factors

- **Deployment-ready:** Designed for real-world implementation with practical constraints

- **Iterative:** Plan for co-creation and refinement with stakeholders

- **Explainable:** Can be understood and trusted by non-technical decision-makers

Goal: Reflect on how a technical proposal could be presented to decision-makers in your field, balancing technical sophistication with practical implementation considerations. Include a system diagram showing the integration of different data modalities and AI components.