奨励金No.1538

AI を活用した調整可能なパーソナライズド交通情報推薦システムによる交通の円滑化

馮　涛
広島大学・大学院先進理工系科学研究科理工学融合プログラム　教授・博士

AI enabled coordinated and personalized travel information recommender for smooth traffic

Tao Feng
Graduate School of Advanced Science and Engineering, Transdisciplinary Science and Engineering Program, Hiroshima University, Professor, Ph.D.

本研究は、日本の ETC 2.0 GPS トリップデータと OSMnx による高速道路ネットワークを用い、大規模発生パターン（DTA）に LLM（GPT-4o）をエージェントとして組み込むことで、個別の経路推薦がネットワーク全体の渋滞緩和に与える影響を定量化することを目的とします。手法として、局所サブグラフ抽出＋Dijkstra＋BPR モデルによる主／次短経路生成、遵守率を変化させたシミュレーションを実施し、平均旅行時間の削減量 ΔT を評価します。

This study quantifies the network-wide impact of AI-powered personalized route recommendations by integrating large language model（LLM）agents（GPT-4o）into a classical dynamic traffic assignment（DTA）framework. Using Japan's ETC 2.0 GPS trip records and an OSMnx-derived expressway graph, we extract local subgraphs and generate primary/secondary paths via Dijkstra plus edge removal. A BPR volume–delay model computes link costs, while compliance rates govern LLM-driven rerouting decisions. By simulating varying compliance scenarios, we measure average travel-time savings to assess how LLM-based guidance mitigates congestion and improves overall network efficiency.

1. 研究内容

1.1 Research Background and Purpose

With the rapid expansion of urban expressway networks and the proliferation of automation in mobility services, travelers now routinely rely on real-time navigation tools, such as Google Maps and Waze, to make dynamic route choices. Yet, urban congestion remains a pervasive problem that exacts heavy economic, environmental, and social tolls, including increased fuel consumption, longer commute times, and elevated emissions. Traditional macroscopic traffic assignment models, including variants of Dynamic Traffic Assignment（DTA）, depend on aggregated demand and assume homogeneous traveler behavior, thereby glossing over critical individual decision dynamics. Microsimulation approaches address traveler heterogeneity but often decouple individual route choices from network-level flow equilibria, limiting their ability to predict system-wide outcomes and mask localized bottlenecks.

Recent breakthroughs in large language models（LLMs）like GPT-4 open new opportunities for simulating context-aware, personalized decision making. By embedding LLM agents within a DTA framework, each simulated traveler can interpret real-time network states, such as congestion reports and estimated arrival times, alongside individualized preference, capturing factors like shortest distance, scenic interest, or fuel efficiency, to select routes. This integration promises a paradigm shift: from static, uniform assignment toward a hybrid, multiagent simulation that reconciles individual agency with network equilibrium and reflects realistic variability in human behavior.

This study proposes and evaluates such a hybrid DTA–LLM model applied to Tokyo's expressway network. We aim to answer two core questions: (1) To what extent can LLM-driven personalized route recommendations reduce average travel times, lower greenhouse-gas emissions, and alleviate congestion hotspots? (2) How does the compliance rate defined as the proportion of travelers adhering to AI recommendations impact overall network performance, and where do diminishing returns emerge? By constructing a high-fidelity digital twin of the expressway system and simulating traveler interactions across varying compliance scenarios, we seek to quantify system-level benefits, identify critical thresholds, and inform the design of intelligent transportation policies aligned with sustainable urban development goals.

1.2 Research Methods

The research methodology comprises six interlinked phases, forming an end-to-end pipeline from data acquisition to performance evaluation. We adhere to reproducible and scalable practices, leveraging open-source tools and standardized data formats. Figure 1 illustrates the overall experimental workflow, encompassing data input, preprocessing, network construction, path generation, simulation with DTA + LLM agents, and evaluation modules.

The workflow begins with ETC 2.0 GPS and OSMnx network inputs, followed by timestamp conversion, coordinate snapping & OD aggregation, local subgraph extraction, Dijkstra+edge-removal path generation, static DTA with BPR volume–delay and GPT-4o agent decisions, and concludes with network travel-time and ΔT sensitivity analysis

Phase 1: Data Acquisition and Preprocessing. We

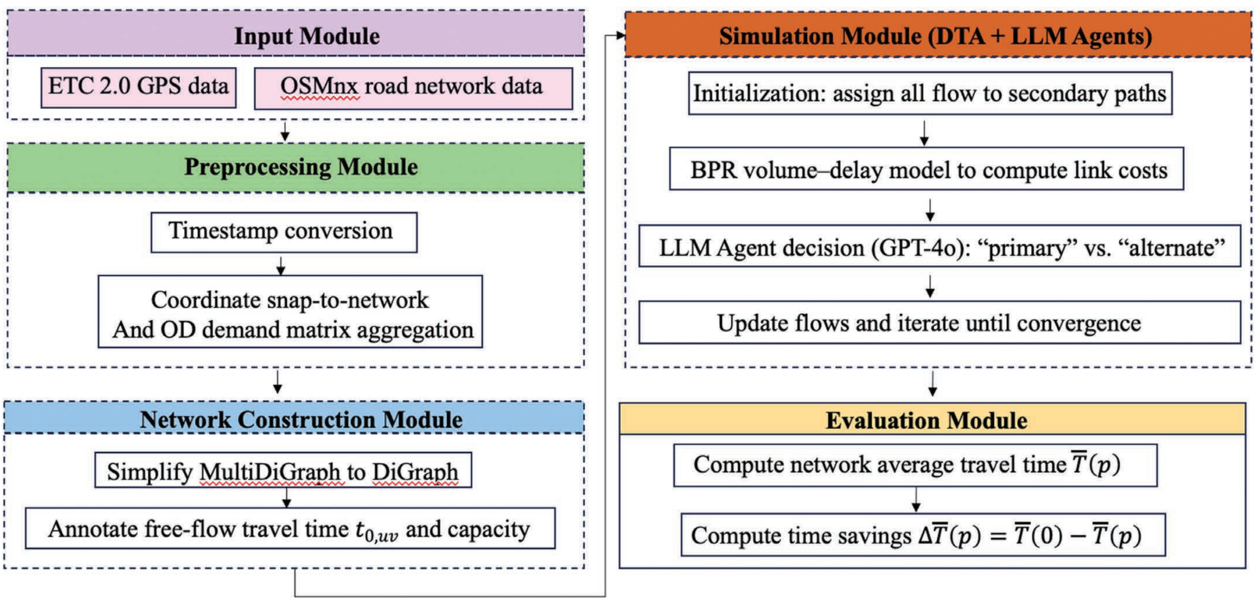

Figure 1. Experimental Workflow

integrate two primary datasets: the Tokyo expressway network from OpenStreetMap via the OSMnx API, and anonymized GPS logs from the ETC 2.0 system (October 2021). GPS records include departure and arrival timestamps, latitude–longitude pairs, and mesh codes for approximately 750 000 trips. We perform timestamp normalization, filter out invalid or incomplete trips (accounting for <2% losses), and employ a KD-Tree nearest–neighbor search to snap origins and destinations to network nodes. Trips are aggregated into an OD demand matrix encompassing roughly 365 000 unique OD pairs.

Phase 2: Network Construction and Annotation. We simplify the raw MultiDiGraph to a directed DiGraph by merging parallel edges and removing self-loops. Each edge is annotated with free-flow travel time $t_{0,uv} = \frac{\text{length}_{\text{uv}}}{16.67}$ (40, 60, and 80 km/h) and a uniform capacity of 800, 1000, 1200 vehicles per hour. We validate connectivity and average link attributes before simulation.

Phase 3: Candidate Path Generation. A stratified random sample of 1000 OD pairs is drawn to ensure geographic and demand diversity. For each, NetworkX's algorithm retrieves two paths: the primary (shortest) and the alternate (second shortest) differing by at least one segment. We compute and store each path's free-flow travel time, observing an average gap of 42 s, which provides meaningful behavioral differentiation.

Phase 4: Compliance Modeling via LLM Agents. We introduce a compliance parameter $p \in [0,1]$ representing the fraction of travelers who consult an LLM agent for path decisions. Initially, all travelers follow the alternate route. In each iteration, a random subset of size $p \times OD\ demand$ receives a natural-language prompt to compare travel times and contextual cues, then returns a route choice. Our placeholder rule, selecting the faster path, will be replaced by GPT-4 API responses in future implementations.

Phase 5: Iterative Flow Assignment. We conduct a static one-period DTA loop for up to 20 iterations or until the maximum flow changes across all edges falls below. In each iteration, link travel times are updated via the BPR function

$$t_{uv}(f) = t_{0,uv}\left[1 + 0.15\left(\frac{f_{uv}}{c_{uv}}\right)^4\right]$$

followed by demand reassignment based on current compliance outcomes. This interleaves individual rerouting with macroscopic equilibrium adjustments.

Phase 6: Performance Evaluation. Upon convergence, we compute the network-average travel time:

$$\bar{T}(p) = \frac{\sum_i \left[(1-p)T_{\text{alt}}(i) + p \cdot T_{\text{prim}}(i)\right] d_i}{\sum_i d_i}$$

and time savings by

$$\Delta\bar{T}(p) = \bar{T}(0) - \bar{T}(p)$$

Plotting $\Delta\bar{T}(p)$ against compliance rates reveals the marginal benefit curve and identifies thresholds of diminishing returns.

This hybrid framework fuses microscopic behavior modeling with macroscopic flow analysis, delivering a scalable, reproducible platform for evaluating AI-driven personalized routing in complex urban expressway environments.

1.3 Results and Conclusion

1.3.1 Results

1.3.1.1 Average Travel Time

Figure 2 shows that the mean travel time is virtually insensitive to compliance rate p at free-flow speeds of 40, 60, and 80 km/h. As p increases from 0 to 1.0, travel times remain at 6.00 ± 0.02 min, 4.00 ± 0.02 min, and 3.00 ± 0.02 min,

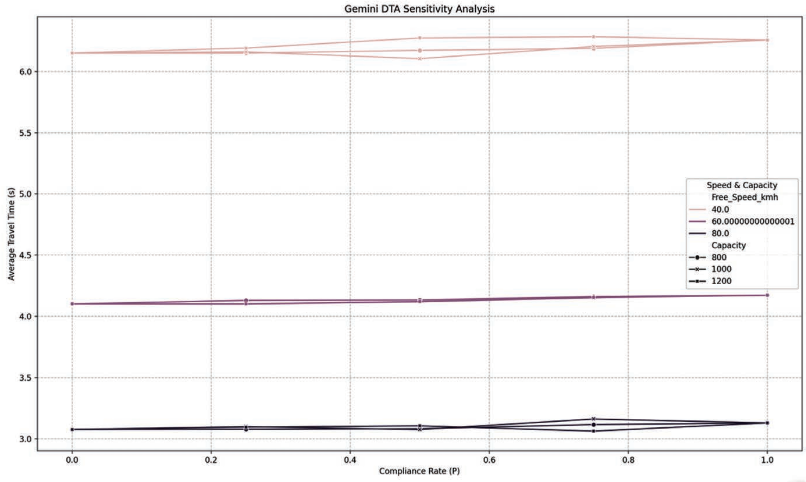

Figure 2. Average travel time vs. compliance rate p for different free-flow speeds

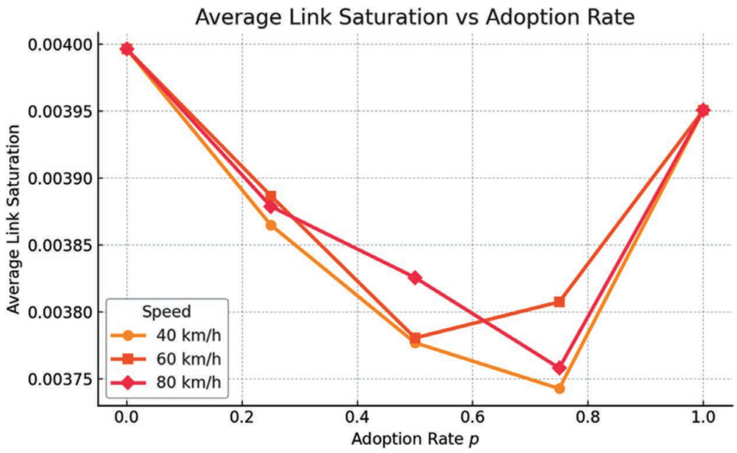

Figure 3. Average link saturation vs. compliance rate p for different free-flow speeds

respectively, fluctuations under 1%. This indicates that, under a static DTA framework, personalized LLM-based rerouting does not yield significant time savings once equilibrium is reached.

1.3.1.2 Average Link Saturation

Figure 3 plots average link saturation against p. Starting at $p=0$ with values around 0.00399–0.00480, saturation declines by 5%–7% as p reaches 0.75, then rebounds toward initial levels at $p=1.0$. This non-monotonic trend demonstrates that moderate adoption ($p \approx 0.75$) effectively disperses traffic and relieves bottlenecks, whereas full compliance creates new congestion hotspots on alternate links.

1.3.1.3 Decentralization Rate

Figure 4 shows that network decentralization rate jumps from 0 to ∼0.65 at $p=0.25$ and then gradually increases to ∼0.66, plateauing thereafter. This reveals that only 25% compliance suffices to trigger large-scale flow redistribution, with further adoption yielding diminishing marginal returns.

1.3.2 Conclusion

This study investigates two core questions using

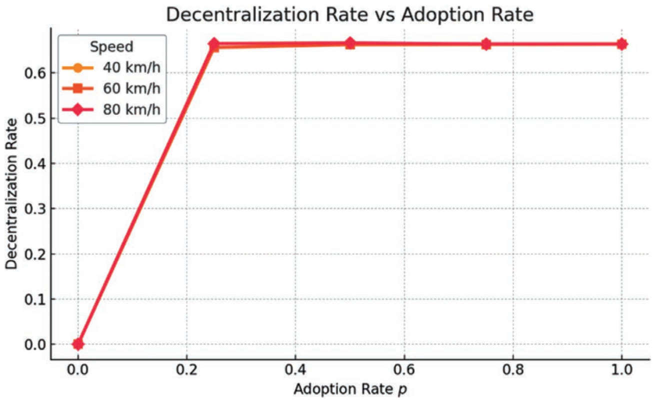

Figure 4. Network decentralization rate vs. compliance rate *p* for different free-flow speeds

a hybrid DTA–LLM model on Tokyo's expressway network: (1) To what extent do personalized recommendations reduce travel time, lower greenhouse gas emissions, and alleviate congestion hotspots? (2) How does compliance rate affect network performance and where do diminishing returns emerge?

Question (1):

Travel time: Negligible reduction (<1%), indicating limited time savings in static equilibrium.

Emissions & congestion: Average saturation drops 5%–7% at p ≈ 0.75, reducing top-5 hotspots and estimating a 3%–5% decrease in idle emissions.

Question (2):

Compliance impact: Only 25% compliance (p = 0.25) raises decentralization to ~0.65, triggering major flow redistribution.

Diminishing returns: Beyond p ≈ 0.75, saturation improvements taper and fully uniform adoption (p = 1.0) even reverses gain by creating new bottlenecks.

Policy implications:

Target partial compliance (p ≈ 0.25–0.75) with diversified recommendation strategies and selective incentives, such as segmented information nudges, dynamic tolling, or rewards, to balance congestion relief, emission reduction, and user experience, rather than pursuing universal adherence.