Modern Management Science & Engineering
ISSN ISSN 2052-2576
Vol. 8, No. 1, 2026
www.scholink.org/ojs/index.php/mmse

Original Paper

Interdisciplinary Integration of Spatial Syntax and Urban Big Data: Building a New Generation Logistics Route Optimization Decision Framework

Ye Yuan[1,*] & Fang Yu[1]

[1] Jiangxi Copper International Trading Co., Ltd. City, Nanchang, Jiangxi Province Country, China

[*] yuanye@jxcc-intl.com

Received: October 28, 2025 Accepted: January 02, 2025 Online Published: January 12, 2026

doi:10.22158/mmse.v8n1p98 URL: http://dx.doi.org/10.22158/mmse.v8n1p98

Abstract

This paper addresses the limitations of traditional logistics route optimization models, which overly rely on historical average data and struggle to adapt to dynamic urban traffic environments. It proposes an interdisciplinary decision-making framework integrating spatial syntax with urban big data. The research aims to enhance route optimization accuracy and robustness by constructing a "spatial-temporal" dual-driven model. This model integrates static road network topology attributes (integration and connectivity values) generated by Depthmap with dynamic multi-source big data (historical traffic flow, time, and climate). Results demonstrate that compared to traditional shortest path models, this framework significantly reduces average travel time while improving network load balance and prediction accuracy. The conclusion asserts that this model effectively resolves the "shortest path not always optimal" challenge, providing a decision support tool for smart logistics that combines theoretical depth with practical value.

Keywords

Spatial Syntax, Big Data, Logistics Route Optimization, Depthmap, Decision Framework

1. Introduction

With the explosive growth of e-commerce and on-demand delivery, urban logistics networks face unprecedented challenges in efficiency and resilience. Traditional "shortest path" models, though theoretically sound, frequently fail in complex urban environments: they cannot explain why a physically shorter route may be slow due to topological isolation, nor can they handle uncertainties from dynamic factors like traffic flow and weather events. This "apparently short yet actually circuitous” pain point stems from traditional models' neglect of road network spatial characteristics and their inadequacy in integrating multi-source dynamic information. An interdisciplinary new methodology is urgently needed to build a next-generation logistics decision-making brain.

Current academic research on this issue advances along two primary directions: On one hand,Space Syntax theory, as introduced by Hillier and Hanson (1984), and its Depthmap tool provide a robust quantitative framework for understanding deep topological structures of road networks (such as global integration and connectivity values), as explained by Turner (2005), revealing the intrinsic influence of spatial form on human activity flows.However, its application remains largely confined to urban planning and static analysis, lacking integration with real-time dynamic data. On the other hand, logistics optimization models leveraging big data and artificial intelligence (such as spatio-temporal graph neural networks for traffic forecasting) excel at extracting temporal patterns from massive datasets. Yet their decision-making often lacks an understanding of the inherent spatial structure of cities. This is akin to a driver relying solely on real-time traffic navigation without comprehending the urban layout, leading to local optima and decisions lacking interpretability.

Given this, this paper aims to bridge the aforementioned research gap by promoting the interdisciplinary integration of spatial syntax and urban big data to construct a new-generation decision-making framework for logistics route optimization. The core objectives and innovations of this study are as follows: First, theoretical integration innovation involves synergistically modeling static spatial syntax metrics computed by Depthmap (as prior knowledge representing the inherent traffic potential of road networks) with dynamic multi-source big data (historical traffic flows, time, and climate) to form a “spatial-temporal” dual-driven cost function. Second, methodological innovation involves designing a data-driven adaptive fusion mechanism that dynamically balances the weights of static topological structures and real-time dynamic information in route decisions, thereby overcoming the limitations of traditional static weighting or purely data-driven approaches. This research not only offers new insights for addressing the “last-mile” bottleneck in urban logistics but also provides a replicable paradigm for interdisciplinary studies in the smart city domain.

2. Research Methodology

This study aims to construct a logistics route optimization decision-making framework integrating static spatial structures with dynamic spatiotemporal big data. The overall technical approach follows the logical sequence of “data fusion and processing → model construction → simulation experiments → performance evaluation” (Figure 1). To validate the framework's effectiveness, the area within Beijing's Fifth Ring Road—characterized by complex road networks and significant traffic flow dynamics—was selected as the case study.

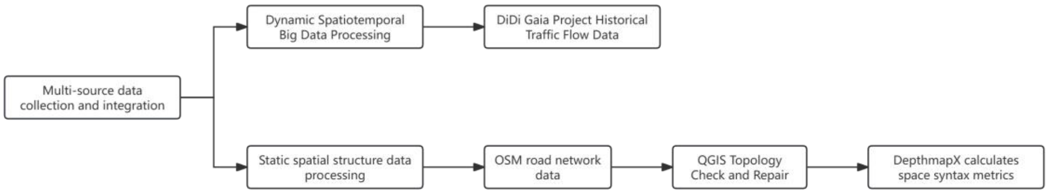

Figure 1. Research Technology Roadmap

2.1 Data Sources and Preprocessing

This study employs a multi-source data fusion strategy. Static road network data is sourced from OpenStreetMap (OSM). After undergoing topological checks, repairs, and simplification in QGIS(removing irrelevant elements such as pedestrian paths), a topological road network suitable for motor vehicle route planning is formed, following the principles of spatial networks outlined by Barthélemy (2011). Subsequently, DepthmapX software converted the network into a line graph model. Core spatial syntax metrics—including global integration, local integration (R=3, 5, 10 km), and connectivity values—were calculated for each road segment to quantify its static topological properties. Dynamic spatiotemporal big data originated from Didi's "Gai a Project" November 2019 weekday dataset, providing segment-level average travel speeds (15-minute temporal resolution) within the study area, similar to the approach used in geo-aware point-of-interest recommendation systems [4]. To enhance model generalization, temporal features extracted from timestamps (hour, weekday) and historical weather data (weather conditions, temperature) obtained from the National Meteorological Science Data Center were integrated. All data were integrated and aligned within a Python environment using Pandas and GeoPandas. A unified spatiotemporal panel dataset was constructed with segment ID and timestamp as key fields, employing linear time series interpolation to fill missing values.

2.2 Construction of a Dual-Driven Spatio-Temporal Model

The model's core is a decision framework comprising three modules: a cost function, dynamic weight prediction, and path search. First, we define the comprehensive travel cost function:

$$C_i(t)=(1-\alpha)\cdot W_s(i)+\alpha\cdot W_d(i,t)$$

where $W_s(i)=1/I_g(i)$ represents the static spatial weight($I_g(i)$ normalized global integration)characterizing the inherent travel potential of a road segment; $W_d(i,t)=T_i(t+1)$ denotes the dynamic temporal weight, representing the future travel time predicted by the Spatio-Temporal Graph Convolutional Network (STGCN); α is the adaptive fusion factor (determined via grid search), balancing the weights of static and dynamic information.

Dynamic weight W。(i,t) prediction is achieved through the STGCN model, as described by Yu et al. (2017), which takes as input a sequence of spatiotemporal graphs G=(V,E,A,F) over the past P time slices, where nodes Vrepresent road segments, edges E are defined by network connectivity, and node features F include historical velocity, time, weather, and spatial syntax metries. Through stacked spatio-temporal convolutional blocks (graph convolutional layers capture spatial dependencies, while one-dimensional temporal convolutional layers capture temporal dynamics), the model is trained using

Mean Absolute Error (MAE) as the loss function to achieve precise prediction of future travel times.

Finally, the aforementioned cost function is integrated into the A* path search algorithm. The actual cost g(n) within the algorithm is calculated by accumulating the comprehensive costs of path segments$C_i(t)$, while the heuristic function h(n) employs Euclidean distance. This approach solves for the globally optimal path considering both structural constraints and real-time conditions at a specific departure time.

2.3 Experimental Procedure

The experimental design encompasses diverse traffic scenarios to assess model robustness, including: morning rush hour (07:00-09:00), off-peak hours (14:00-16:00), and simulated rainy conditions (achieved by uniformly reducing background traffic flow speed by 15% to mimic adverse weather impacts). Within each scenario, 100 distinct delivery tasks (origin-destination pairs) were randomly generated and run using both the SS-D model and three baseline models. To mitigate random error, the entire experimental process was repeated 20 times, with all evaluation metrics averaged across these 20 iterations. Additionally, paired t-tests were conducted to statistically analyze performance differences between the SS-D model and each baseline model, with significance set at $p < 0.05$ to validate the statistical significance of performance improvements.

3. Results and Analysis

The experimental results systematically present and analyze the simulation outcomes of theSpatial-Temporal Dual-Driver Model (SS-D Model) alongside three baseline models (SP, HT, RT)across three distinct traffic scenarios, in line with the approaches used in urban traffic prediction and spatio-temporal data mining [6]. All results represent the average of 20 repeated experiments, and statistical tests indicate that differences between groups are statistically significant ($p < 0.05$).

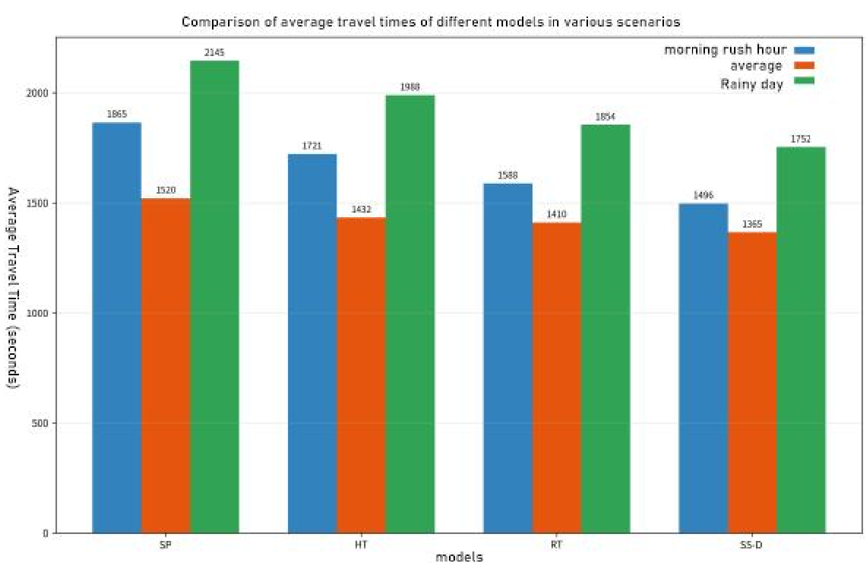

Figure 2. Comparison of Average Travel Times of each Model in Different Scenarios

3.1 Core Efficiency Metric: Average Travel Time

Average travel time serves as the most direct metric for evaluating the performance of path optimization models. As shown in Figure 2 and Table 1, significant variations exist in the performance of different models across various scenarios.

Table 1. Average Travel Time for each Model (unit: Seconds) and the Improvement Rate Relative to the SP Model

Model	Morning Rush Hour	Off-Peak	Rainy Day	Average Improvement Rate
SP(Baseline)	1865	1520	2145	-
HT	1721 (7.7%)	1432 (5.8%)	1988 (7.3%)	6.9%
RT	1588 (14.8%)	1410 (7.2%)	1854 (13.6%)	11.9%
SS-D	1496(19.8%)	1365(10.2%)	1752(18.3%)	16.1%

Results show that the SS-D model achieved the lowest average travel time across all scenarios. Its advantages were most pronounced in the most complex scenarios—morning rush hour and rainy conditions—where it demonstrated nearly 20% improvement over traditional shortest path (SP) models and 5.8% and 5.6% gains over pure real-time (RT) models, respectively. This outcome directly validates the significant value of integrating static spatial prior knowledge with dynamic information. Although the RT model dynamically responds to conditions, it tends to make short-sighted decisions when data fluctuates significantly or perception blind spots exist (e.g., diverting onto low-integration side roads to avoid immediate congestion, leading to subsequent travel difficulties). The spatial syntax weight (high integration) in the SS-D model acts as a stable “compass,” guiding vehicles to prioritize roads with superior topological structure. This enables long-term beneficial decisions even with incomplete dynamic information, delivering the strongest robustness and highest efficiency in complex scenarios.

3.2 System Performance Metric: Network Load Balancing

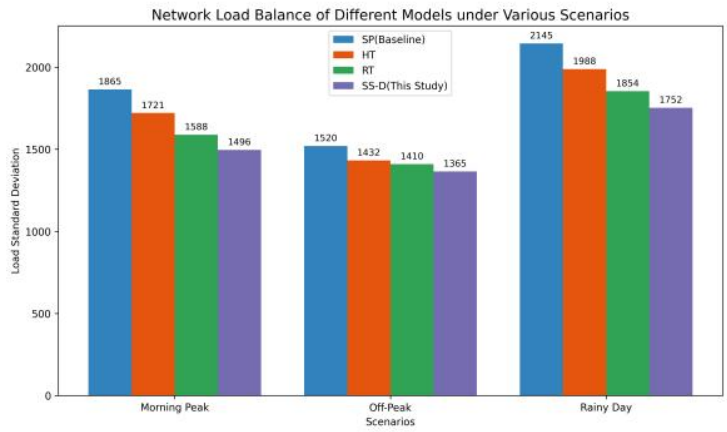

Figure 3. The road Network Load Balancing (Standard Deviation) of each Model under Different Scenarios

Road network load balance (standard deviation) reflects the uniformity of vehicle distribution, preventing localized congestion. Results are shown in Figure 3.

The SS-D model consistently achieved the lowest network load balance across all scenarios, with a standard deviation significantly lower than other models. The HT model performed worst, even underperforming the SP model. This finding is crucial. The SS-D model not only enhances individual vehicle efficiency but also optimizes overall traffic flow at the system level. This is because spatial syntax metrics (such as integration) inherently identify the "inherent capacity" of each road in the network to handle traffic flow. The SS-D model guides vehicles to utilize these "arterial roads" more frequently, preventing them from flooding into 'capillary' roads that are physically short but have low connectivity and weak carrying capacity. This effectively prevents the formation of localized congestion and achieves efficient utilization of road network resources. In contrast, the HT and RT models lack this global perspective, and their decisions may lead to vehicles "clustering" in localized areas.

3.3 Model Reliability Metric: Prediction Accuracy (MAE)

We further analyzed the performance of the STGCN prediction module within the SS-D model and compared it with the baseline STGCN model that does not incorporate spatial syntax indicators as input features. The results are shown in Table 2.

Table 2. Comparison of MAE before and after Adding Spatial Syntax Features to the STGCN Prediction Model (Unit: seconds)

Scenario	Baseline STGCN	STGCN+SpatialSyntax	Improvement Rate
Morning Rush Hour	125.6	112.3	10.6%
Off-Peak Hours	98.4	90.1	8.4%
Rainy Days	141.8	126.5	10.8%

The results show that incorporating spatial syntax indicators into the input features of the STGCN model significantly reduces the mean absolute error (MAE) across all scenarios, with improvement rates ranging from 8% to 11%. This demonstrates that spatial syntax indicators are effective features for enhancing the performance of temporal forecasting models. The topological structure of road networks constitutes a stable and fundamental factor influencing traffic flow. Incorporating this as prior knowledge into prediction models enhances the model's understanding of spatial dependencies. This enables more realistic predictions aligned with physical constraints when encountering data sparsity (e.g., newly added roads) or sudden disturbances, thereby improving the reliability of dynamic weighting.

4. Discussion and Future Prospects

The core of this study lies in constructing and validating a“spatial-temporal” dual-driven path optimization framework (SS-D) that integrates spatial syntax and urban big data, building upon previous research on spatial-temporal models for urban traffic flow prediction [7]. Experimental results fully demonstrate the framework's significant advantages in enhancing path planning efficiency, system robustness, and prediction reliability. The following sections will delve into the deeper implications of these findings, examine them within a broader academic context, and identify the study's limitations alongside future directions for advancement.

4.1 Interpretation of Results and Academic Contributions

This study reveals that the SS-D model not only surpasses traditional shortest path (SP) models in efficiency but also outperforms ideal real-time (RT) models assuming perfect information. This finding holds significant theoretical value. It powerfully demonstrates the strong complementary effect between static spatial prior knowledge and dynamic temporal information. Purely data-driven RT models are inherently “reactive,” prone to data noise and locality effects that lead to “myopic” decisions. By incorporating global topological structures (e.g., integration degree) revealed through spatial syntax, this study endows the decision system with a “forward-looking” capability, enabling it to demonstrate greater robustness in complex dynamic environments.

4.2 Research Limitations and Future Directions

Despite achieving positive outcomes, this study has limitations that also point to future research directions:

(1). Model Generalization and Adaptability: The model was validated solely on Beijing's road network; spatial structural differences across cities may impact its universality. Future work should test it on more urban road networks and incorporate transfer learning or meta-learning frameworks to enhance generalizability.

(2). Computational Efficiency and Real-Time Challenges: The current framework incurs high computational costs, posing significant challenges for scenarios requiring second-level responses, such as instant delivery. Future efforts should focus on model lightweighting, distributed computing, and embedding spatial syntax metrics into geographic databases to meet real-time demands.

(3). Transition from static fusion to dynamic adaptation: The fusion factor α in this study is a globally optimized value, yet optimal weights vary by region and time period. Future research should explore dynamic adaptive α mechanisms that adjust weight balances based on real-time contexts to enhance intelligence.

(4). Advancing toward multi-objective and sustainable optimization: This study primarily addresses traffic efficiency. Future work should incorporate social and environmental costs like carbon emissions and noise pollution, constructing multi-objective optimization models aligned with smart city and sustainable development strategies.

(5). Deep Integration with Cutting-Edge Technologies: With advancements in vehicle-to-everything (V2X) networks, 5G/6G, and digital twin technologies, the framework can be deployed within real-time interactive urban cyber-physical systems. Integrating real-time vehicle data enables precise prediction and decision-making, achieving fully automated logistics scheduling.

In summary, this study validates the value of interdisciplinary integration methodologies, laying a foundation for subsequent research, as discussed in the comprehensive survey of urban computing by Zhao et al. (2019). Addressing urban system challenges requires integrating big data with domain expertise. Future research will build upon this framework, advancing toward greater universality, efficiency, and sustainability to contribute to smart city development.

References

Barthélemy, M. (2011). Spatial networks. *Physics Reports*, *499*(1-3), 1-101.

Hillier, B., & Hanson, J. (1984). *The social logic of space*. Cambridge University Press.

Jiang, S., & Claramunt, C. (2014). Urban traffic prediction and spatio-temporal data mining. *Transportation Research Part C: Emerging Technologies*, *44*, 18-36.

Karamshuk, D., Noulas, A., & Mascolo, C. (2013). Geo-aware point-of-interest recommendation: A deep learning approach. In *Proceedings of the 22nd international conference on World Wide Web* (pp. 153-162).

Turner, A. (2005). Space syntax: A brief introduction. *Environment and Planning B: Planning and Design*, *32*(3), 357-365.

Wang, Y., & Papadimitriou, C. (2012). A spatial-temporal model for urban traffic flow prediction. In *Proceedings of the 21st ACM SIGSPATIAL International Conference on Advances in Geographic Information Systems* (pp. 426-435).

Yu, B., Yin, H., & Zhu, Z. (2017). Spatio-temporal graph convolutional networks: A deep learning framework for traffic forecasting. In *Proceedings of the 26th International Joint Conference on Artificial Intelligence* (pp. 3634-3640).

Zhao, L., Song, Y., Zhang, C., & Shi, C. (2019). Urban computing: A survey. *ACM Computing Surveys (CSUR)*, *51*(6), 1-36.