A regional-scale estimation model for residents' travel structures based on the multi-modal travel data

Huapeng Shen[a]; Jiancheng Weng[a]; Pengfei Lin[b]

a Beijing Key Laboratory of Traffic Engineering, Beijing University of Technology, Beijing, 100124, China

b Faculty of Information Technology, Beijing University of Technology, Beijing 100124, China

Introduction

Driven by the need to reduce greenhouse gas emissions, alleviate traffic congestion, and promote environmental sustainability, the development of low-carbon, green and sustainable transportation system has become a core focus of urban planning and development. While significant progress has been made in sustainable transportation through the implementation of policies such as Transit-Oriented Development (TOD), several challenges remain. At the same time, with the emergence of new travel modes, such as carpooling, shared vehicles, autonomous taxis, and flying cars, there is a growing diversity in travel choices available to different groups, resulting in more complex travel chains for residents. Therefore, understanding the usage of different transportation modes within a region, characterizing the regional travel structure, and analyzing the mechanisms underlying changes in regional travel patterns are essential for formulating targeted policies and interventions. These efforts are crucial for increasing the proportion of green and low-carbon travel and promoting sustainable travel behaviors.

Although a substantial body of research has been conducted on green guidance strategies for residents and their spatial heterogeneity, several issues remain unresolved, Firstly, during the community space planning and transportation system optimization phases, urban planners and managers often focus on the macro-level travel structure proportions of the city, neglecting the heterogeneity of travel structures in localized areas [1]. Secondly, current studies on residents' travel behavior or travel structure are based solely on survey data, utilizing discrete choice models or machine learning for analysis, without capturing the long-term dynamic evolution of residents' travel behavior [2]. To address these issues, this study proposes a regional residents' travel structure evaluation model based on multi-semantic learning methods and Large Language Models (LLM). It provides quantitative support for the discovery of spatiotemporal patterns in regional travel structures, analysis of low-carbon travel structures, and identification of the potential for regional low-carbon travel.

Methodology

The study constructs a regional travel structure estimation model based on multi-modal data, employing multi-semantic learning methods and the LLM to forecast residents' travel patterns across different spatial contexts and assess regional travel structures. The specific process is illustrated in Figure 1. The model primarily involves two methods: (1) regional multi-modal data feature extraction based on multi-semantic learning, and (2) regional travel structure prediction using the LLM model.

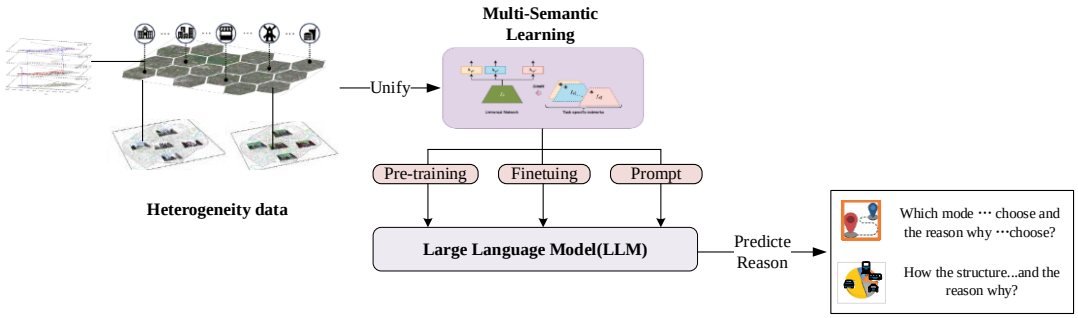

Figure1 The Proposed Framework for Estimating the Regional Travel Structure

Feature extraction of region multi-modal data based on multiple semantic learning

In this section, the study utilizes multi-semantic learning methods to integrate multi-source multi-modal data. To characterize the spatiotemporal characteristics of different regions, Specifically, this study first considers the distinct text and image information associated with each region and proposes a method for extracting the spatiotemporal features of text and images for each region using a pre-trained encoder-based model:

$$e_i^{text} = \frac{1}{|T_i|}\sum_{j=1}^{|T_i|} W(T_{ij}) \tag{1}$$

$$e_i^{image} = \sum_{m\in\{SV,RV\}} \beta_i^m e_i^m \tag{2}$$

where T is the category of features in the region and the final mapping of the trained model from words to vectors is W; e_i^{image} the final representation for region's imagery feature. Subsequently, this study employs a feature-level attention fusion module, which aligns the combined image features with the text representation vectors for each region. This approach injects both visual and textual semantic insights into the fused features:

$$Loss_i = -log\frac{exp\left(sim\left(e_i^{text},e_i^{image}\right)\right)}{\sum_{j=1}^{n} exp\left(sim\left(e_i^{text},e_i^{image}\right)\right)} \tag{3}$$

Estimation model for regional travel structure based on LLM model

Unlike traditional prediction and evaluation models, the LLM model can automatically capture complex semantic patterns through vast amounts of pre-trained data, overcoming the limitations of traditional models that rely on manual feature engineering [3]. Therefore, this study uses LLM to assess regional residents' travel outcomes. In this section, we provide the unified mathematical definition of the LLM, including its inputs, outputs, and objective function. The specific method is illustrated in Figure 2. First, for a given region i, the input to the LLM model δ_i can be represented by the following combination in natural language, based on the design of the prompt:

$$\delta_i = (p, \rho_i) \tag{4}$$

where p represents the prompt; ρ_i represents the multi-modal features extracted for the region. Given these input tokens, this study employs a widely used LLM with millions of

parameters to generate regional travel structure metrics and calculations based on its understanding of specific task instructions:

$$\mu_i = LLM(\delta_i) \tag{5}$$

where μ_i represents the target demand label. Finally, this study reformulates the regional travel structure metric goal as a conditional language generation task, and optimizes the metric goal by minimizing the negative log-likelihood (NLL) of the target labels:

$$L_{NLL} = -\sum_{i=1}^{I}\sum_{k=1}^{K} log P_\theta\left(\mu_i^k \middle| \delta_i, \mu_i^{<k}\right) \tag{6}$$

where I and K represent the numbers of regions and target tokens, P_θ is the probability distribution of tokens based on the model parameters θ.

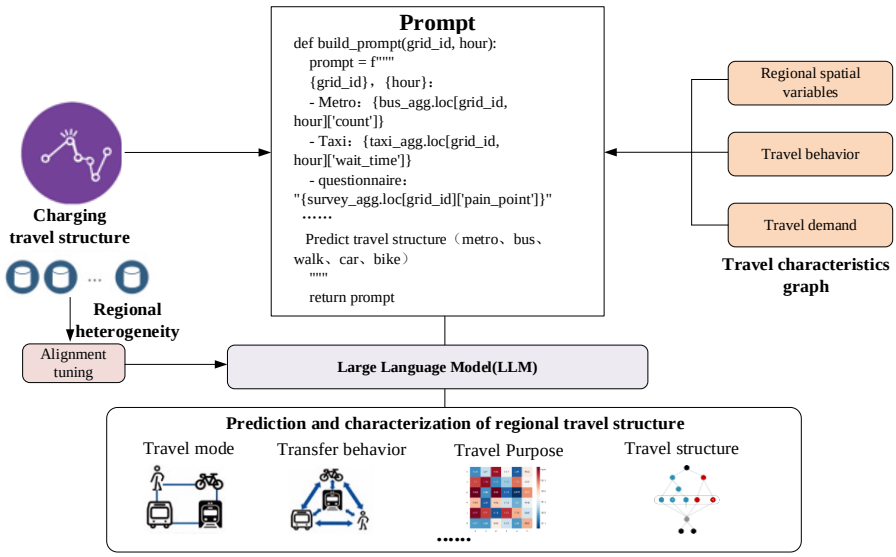

Figure 2 Regional Travel Structure Estimation Framework Based on LLM Model

Results

The study takes Beijing as a case example and utilizes the proposed framework to estimate the travel structures of residents in different spatial communities within the city. The results characterize the travel structure features of various communities and compares the results with traditional predictive models to validate the accuracy of the proposed framework. As shown in Table 1, compared with traditional models, the LLM demonstrates significant advantages in multi-modal fusion, semantic reasoning, and small sample adaptation.

Table 1 Comparison of predictive performance among various models

Model	RMSE	MAE
ARIMA	6.58	4.33
LSTM	6.08	4.01
GCN	5.83	3.46
LLM	5.33	3.23

Furthermore, as shown in Table 2, the results also reveal that the community-level travel structure in Beijing exhibits significant spatial differentiation and temporal dynamics. In the spatial dimension, in the core area (such as ChaoyangCBD), the

coverage of rail transit stations within 500 meters reaches 98%, whereas in peripheral communities (such as Changyang, Fangshan district), it is only 45%, directly leading to differences in private car dependency (15% in core areas vs. 40% in peripheral areas) In terms of the temporal dimension, in commuting communities, metro ridership during peak hours accounts for 65% of daily usage, while in cultural and tourist communities, the use of shared bicycles on weekends surges by 50%.

Table 2 Characterization of travel structures in different communities

Region	Public transport	Car	Active travel	Key features
Zhong guan cun	55%	15%	30%	Subway+shared bicycle connection
Hui long Temple	40%	35%	25%	The full load rate of the subway during the morning peak hour, relying on customized public transport
Lu cheng	34%	42%	24%	Work to residence ratio<0.8, long commuting distance

Conclusion

The analysis of regional residents' travel results and evaluation models indicates that the LLM model demonstrates significant improvements in various aspects compared to traditional models; while, the use of multi-semantic learning methods to extract regional spatial features effectively characterizes spatial differences between regions, addressing the issue of spatial variation that existing studies fail to account for; and finally, the analysis of Beijing reveals spatiotemporal disparities of travel structure, and this disparities influenced by factors such as the gradient of regional transportation infrastructure supply, job-housing balance, the interaction with the built environment, peak hours, and seasonal fluctuations, leading to different travel structures across regions. In the future, authorities can develop targeted strategies for optimizing community-level transportation systems based on the study's findings. For example, for high-density employment communities, micro-circulation feeder buses can be added to increase the share of public transport. However, this study still has some limitations. In the future, more diverse data could be incorporated into the model to improve its accuracy.

References

[1] Næss, Peter. Four common misconceptions in quantitative studies of the built environment and travel. Transportation Research Part D-Transport and Environment, 104597, 2025.

[2] Ghorbani, A., Nassir, N., Lavieri, P.S. et al. Enhanced utility estimation algorithm for discrete choice models in travel demand forecasting. Transportation (2025).

[3] Thilo Hagendorff, Sarah Fabi, and Michal Kosinski. Human-like intuitive behavior and reasoning biases emerged in large language models but disappeared in chatgpt. Nature Computational Science, 3(10):833–838, 2023.