Algorithm for Border Port Personnel Flow Analysis and Economic Vitality Assessment Based on Spatiotemporal Big Data

Ximeng Zhang[1,*], Yuan Yao[2]
[1]International Economics and Trade, School of Business, College of Humanities & Information Changchun University of Technology, Changchun, 130122, China
[2]Party and Government Office, Changchun Technician College, Changchun 130000, Jilin, China
Corresponding author: Ximeng Zhang (e-mail: 18543001880@163.com).

ABSTRACT The conventional techniques utilizing statistical methods and single time series are relatively inadequate to analyze the complex spatiotemporal correlations and economic connections, leading to the challenges in performing refined dynamic analyses of population flow and its influence on the economy. Hence, This article will use a novel fusion approach that integrates the spatiotemporal big data analysis and graph neural network to build a unified model for analyzing the dynamic behavior and economic potential of population flow at border ports. From a methodological perspective, indi vidual mobility patterns are characterized via spatiotemporal alignment and trajectory reconstruction, where this article further create the spatiotemporal graph of flows to describe the interregional migration relations. After that, we design a spatiotemporal graph neural network along with a multi-scale dynamic modeling strategy to predict population flows. As indicated by the experiments, the proposed method performs exceptionally well in population flow prediction tasks, obtaining MSE of 0.071 and MAE of 0.174, which demonstrates significant outperformance compared to the conventional approach. Through this study, we prove the effectiveness of our method in achieving refined population flow analysis and providing reliable data and decision-making support for assessing regional economic vitality.

INDEX TERMS border port population flow analysis, economic vitality assessment, spatiotemporal big data, spatiotemporal graph neural network, multi-scale modeling

I. INTRODUCTION

In light of growing cross-border activities, port borders have emerged as essential nodes for the mobility of people and regional economic interactions. The fluctuations in population mobility scale and structure directly indicate the extent of regional openness while having significant consequences for the surrounding business transactions and consumer behaviors. In this context, it is imperative to explore the potential for improving population flow characterization at ports through spatiotemporal big data and subsequently evaluating their influence on economic activity. This paper proposes a comprehensive analytical framework for analyzing the population mobility and evaluating economic vitality at port borders using spatiotemporal big data. Through spatiotemporal synchronization and trajectory reconstruction of multiple sources of data, a spatiotemporal flow network is created for characterizing the population mobility pattern. Based on the spatiotemporal flow network, a spatiotemporal graph neural network model in combination with multi-scale dynamic modeling strategy is proposed for predicting the population mobility process.

Additionally, a set of population flow and economic vitality assessment system, incorporating factors like flow intensity, time spent, and crossing frequency, is developed, and the dynamic coupling mechanism between population flow and economic vitality is identified using nonlinear mapping and lag causality analysis techniques. The current study, utilizing a comprehensive spatiotemporal modeling and economic assessment framework, offers a novel approach for analyzing population flow and characterizing regional economic activity.

II. RELATED WORK

In the research of population flow and spatial structure field, the scholars have done a lot of systematic explorations from various perspectives. Tao et al. studied the spatial relationship and population flow pattern of urban area according to the migration of population, indicating that the relationship of urban space has a “diamond” structure and the population flow in the Spring Festival holiday has symmetry and follows the distance decay rule [1], which gives us a good reference about the macro distribution characteristic of population flow. On the other hand, Jiang et al. began with the perspective of spatial behavior at a micro level, establishing an indoor evacuation optimum path model based on Dijkstra's algorithm and realizing the dynamic path design of population flow by applying computer vision and BIM techniques [2]. In terms of economic behavior, Liu et al. quantitatively analyzed the spatial distribution feature of corporative algorithmic collusion, indicating that it tends to cluster in the eastern and central regions and it is significantly associated with the level of internet facilities and human resources [3];Huang focused on the usage of artificial intel ligence in the area of human resources management, stressing the significance of using technology, data, and policy interaction in optimizing the system [4]. Furthermore, Wang et al. optimized the Grounding DINO model for traffic passenger flow detection by implementing the multimodal fusion and lightweight method, leading to high detection accuracy [5]. Moreover, Wang et al.'s other research investigated the flow patterns of scientific researchers via complex networks and discovered its features as a scale-free network and regional differences [6].In conclusion, previous research has produced fruitful findings in capturing the patterns of macro-level migration, optimizing the path at the micro-level, network models, and multilayer perception. Nevertheless, the previous research work mostly concentrated on a specific scenario or a single approach but did not address the interaction between "spatiotemporal flows and economic vitality."

III. METHODS

A. SPATIOTEMPORAL DATA MODELING AND FLOW NETWORK CONSTRUCTION

1) Multi-Source Data Fusion and Spatiotemporal Alignment Processing

In border crossing population flow analysis, data typically originates from various heterogeneous systems such as mobile signaling, customs clearance records, and traffic checkpoints. Since different data sources differ in temporal granularity and spatial precision, unified spatiotemporal alignment and standardization are essential as the first step.

In the time dimension, a unified time window ΔT is introduced to (t_i) map the original timestamps to discrete time indices:

$$t' = \left\lfloor \frac{t_i}{\Delta T} \right\rfloor \tag{1}$$

This method enables the alignment of multi-source data on the same time scale, thereby eliminating the impact of differences in sampling frequency.

In the spatial dimension, continuous geographic coordinates are $((x_i, y_i))$ mapped to discrete raster cells (r_j), i.e.: $g(x_i, y_i) \to r_j$. This constructs a unified spatial representation system, enabling data from different sources to be compared and integrated within the same spatial grid.

After processing, each record is uniformly represented $((u_i, r_j, t', a_i))$, thus forming a standardized spa tiotemporal data structure, providing a foundation for subsequent trajectory modeling.

2) Spatiotemporal Trajectory Reconstruction and Feature Representation

After data alignment, trajectory reconstruction is required for individual discrete observation points to recover their continuous movement paths within the port area. For an individual (u), its original observation sequence can be represented as:

$$S_u = (r_1, t_1), (r_2, t_2), \ldots, (r_n, t_n) \tag{2}$$

Because the actual data has sampling discontinuity issues, it is necessary to use interpolation methods to complete the trajectory.

When the time interval between adjacent observation points is large, linear interpolation is used to recon struct the path:

$$\hat{r}(t) = r_i + \frac{t - t_i}{t_{i+1} - t_i}(r_{i+1} - r_i) \tag{3}$$

This results in a continuous trajectory $(\widetilde{\mathcal{T}}_u)$, making the individual's movement process smoother in the time dimension.

Building upon this, a stop-point identification mechanism is introduced to characterize individual behavioral states. When the individual's movement speed satisfies:

$$v_i = \frac{\mathrm{dist}(r_i, r_{i+1})}{t_{i+1} - t_i} < \epsilon_v \tag{4}$$

If the dwell time meets the requirements $(t_{i+1} - t_i) > \epsilon_t$, the location is determined to be a valid dwell point, thereby identifying the individual's dwelling behavior pattern at the port. Further, in the feature expression stage, temporal periodicity features and spatial distribution features are introduced. Temporal features are used to characterize the periodic changes in pedestrian flow, for example: $f_t = [\sin(2\pi t/24), \cos(2\pi t/24)]$. Spatial features are encoded based on the grid they reside in, used to describe individual spatial preferences. Ultimately, each individual's trajectory is represented as a feature set that integrates temporal, spatial, and behavioral states.

3) Construction of Port Spatiotemporal Flow Network

Building upon trajectory modeling, the port area is further abstracted into a spatiotemporal flow network to characterize the migration relationships between different functional areas. The port flow network is defined as follows:

$$G = (V, E, W) \tag{5}$$

Where (V) represents the set of spatial nodes, (E) represents the set of personnel flow edges, and (W) represents the edge weight.

When an individual migrates from node (i) to node (j), a directed edge is formed (e_{ij}), and the edge weight is determined by accumulating the number of migrations. The calculation method is as follows:

$$w_{ij} = \sum_t \mathbb{I}(r_t = i \wedge r_{t+1} = j) \tag{6}$$

This indicator reflects the intensity of personnel flow from region (i) to region (j) per unit time, thus charac terizing the main passage routes within the port.

Furthermore, to reflect the dynamic changes in the flow, the network is extended to a time-evolutionary form:

$$G_t = (V, E_t, W_t) \tag{7}$$

The edge weights are updated over time and can be further normalized to flow probabilities $P_{ij}(t) = \frac{w_{ij}(t)}{\sum_k w_{ik}(t)}$.

B. SPATIOTEMPORAL FLOW MODELING AND HUMAN BEHAVIOR PREDICTION

1) Spatiotemporal Feature Representation and Density Modeling

A kernel density estimation (KDE)-based method is introduced to model the spatial distribution of people. At any spatial location (x), its density function can be expressed as:

$$\rho(x) = \frac{1}{nh} \sum_{i=1}^{n} K\left(\frac{x - x_i}{h}\right) \tag{8}$$

Where $(K(\cdot))$ is the kernel function, and (h) is the bandwidth parameter, (x_i) representing the sample location point. In practical applications, spatial regions are usually discretized into grid cells; therefore, grid statistical methods can also be used to calculate the pedestrian flow intensity within a unit area.

$$D_j(t) = \frac{N_j(t)}{A_j} \tag{9}$$

The number of people who fall into grid (j) within time (t) is represented (A_j) by $(N_j(t))$ the area of that grid.

2) Spatiotemporal Graph Neural Network Prediction Model

After obtaining the spatiotemporal feature representation, a spatiotemporal graph neural network (ST-GNN) is introduced to predict personnel flow in order to characterize the complex spatial dependencies between port areas. The port area is abstracted as a graph structure $(G = (V, E))$, where nodes represent spatial units and edges represent personnel flow relationships.

In the spatial modeling section, neighborhood information is aggregated through graph convolution opera tions, and its basic form is as follows:

$$H^{(l+1)} = \sigma\left(\widetilde{D}^{-1/2}\widetilde{A}\widetilde{D}^{-1/2}H^{(l)}W^{(l)}\right) \tag{10}$$

Where is $(\widetilde{A} = A + I)$ the adjacency matrix with self-loops, $(\widetilde{D})$ is the degree matrix, $(H^{(l)})$ and represents the node representation of the (l)th layer.

In the time-based modeling section, a sequence model is introduced to model flow changes, such as a cyclic structure or a gating mechanism. Its state update process can be represented as:

$$h_t = f(h_{t-1}, x_t) \tag{11}$$

Here, (x_t) represents the current flow feature input, (h_t) and represents the hidden state, used to charac terize the dynamic evolution process over time.

By jointly modeling spatial graph structures and time series models, dynamic prediction of port passenger flow can be achieved, namely:

$$\hat{Y}_{t+1} = F(G, X_{1:t}) \tag{12}$$

This $(X_{1:t})$ represents the historical spatiotemporal observation sequence.

3) Multi-Scale Dynamic Modeling Mechanism

Since the migration behavior at the port of entry follows an obvious pattern of periodicity and abruptness, it becomes impossible to use a single-time scale for modeling. Thus, a multi-scale modeling approach is proposed that helps in the analysis of short-run and long-run dynamics separately. First, the input sequence is decomposed into short-term and long-term components:

$$X_t = X_t^{(s)} + X_t^{(l)} \tag{13}$$

This $(X_t^{(s)})$ indicates short-term fluctuations, mainly reflecting hourly or intraday changes; $(X_t^{(l)})$ it also indi cates long-term trends, depicting weekly or monthly changes.

During the prediction process, the two types of features are modeled separately and then fused:

$$\hat{Y}_t = \alpha f_s(X_t^{(s)}) + (1 - \alpha) f_l(X_t^{(l)}) \tag{14}$$

Where $(f_s(\cdot))$ and $(f_l(\cdot))$ represent the short-term and long-term prediction functions, respectively, and (α) is the adaptive weight parameter.

C. ECONOMIC VITALITY ASSESSMENT AND POPULATION FLOW COUPLING ANALYSIS METHOD

1) Construction and Integration of Economic Vitality Indicators

Let the intensity of population movement be (F_t), the average length of stay be (S_t), and the frequency of cross-border travel be (C_t).To avoid the common-sense error of treating all spatial units equally, we introduce a functional zone coefficient (η_j) based on the economic attributes of the grid cell (e.g., high weights for retail/duty-free zones and low weights for administrative transit corridors). Then, the basic economic vitality for a specific zone can be expressed as a weighted combination:

$$E_t = \eta_j(\alpha F_t + \beta S_t + \gamma C_t) \tag{15}$$

Here, (α, β, γ) represents the weighting coefficient, used to balance the contribution of different indicators to economic vitality.

To prevent the misinterpretation of logistical bottlenecks or congestion—which often manifest as extended dwell times—as positive economic indicators, this framework integrates a "congestion-adjusted" utility factor. Based on the above, consumer behavior measures, spatial clustering traits, and a penalty function for non-productive idle time are further introduced into the assessment to enrich the profile of the economic vitality indicator, ensuring that the model distinguishes between productive economic engagement and mere physical presence caused by processing delays. Consumer behavior measure is the size of transactions made by consumers and their purchasing frequency, whereas spatial clustering is the level of clustering of individuals in a certain space, which may be calculated through entropy or density measures. The overall economic vitality indicator may then be extended to:

$$E_t^* = E_t + \lambda_1 M_t + \lambda_2 G_t \tag{16}$$

Wherein (M_t) represents the intensity of consumption behavior, (G_t) represents the degree of spatial agglomeration, (λ_1, λ_2) and is the adjustment coefficient.

2) Spatiotemporal Coupling Modeling

Building upon the assessment of economic vitality, further analysis is needed to determine the coupling relationship between population mobility and economic vitality. Since the relationship between the two is typically non-linear, a function mapping approach is employed for modeling.

Let the characteristics of population mobility be denoted as (X_t) and economic vitality be denoted as E_t^*, then the mapping relationship between the two can be expressed as:

$$E_t^* = f(X_t) \tag{17}$$

Here, $(f(\cdot))$ a nonlinear mapping function is represented, which can be implemented by a regression model, neural network, or graph learning model. To enhance the expressive power of complex relationships, a multivariate joint modeling approach is introduced $E_t^* = f(F_t, S_t, C_t, M_t, G_t)$.

3) Causal Lag Effect Analysis and State Characterization

As economic activities tend to have time lags for changes in population flow, it would be inadequate to depend only on simultaneous relationships to describe their dynamics properly. Hence, time lagging is incorporated in modeling causal relationships. The selection of the lag order (τ) is determined by the specific economic sector being analyzed: short-term lags (τ_{short}) are used to capture immediate retail consumption, while long-term lags (τ_{long}) are applied to model delayed impacts on logistics and trade services.

Let the lag order be (τ), then the relationship between economic vitality and historical population flow can be expressed as $E_t^* = \sum_{\tau=1}^{T} \theta_\tau X_{t-\tau}$, where (θ_τ) represents the influence weight under different time lags, used to measure the contribution of historical population flow to current economic vitality.

Based on this, we can further analyze the causal directionality and determine the dominant relationship of "people flow to economy" or "economy to people flow" by comparing the intensity of the lagged effects, thereby characterizing the causal structure of the system. While this lag analysis identifies temporal prece dence, it is recognized as a "constrained causality" within the observed data. To enhance robustness against exogenous shocks, the model incorporates seasonal adjustment filters and dummy variables for major policy shifts, ensuring that the identified coupling reflects the intrinsic interaction between mobility and economy rather than coincidental alignment driven by external trade cycles.

Finally, to provide a structured description of the port's economic operation, a state partitioning function can be constructed based on the economic vitality index, for example:

$$S_t = g(E_t^*) \tag{18}$$

represents (S_t) the category of economic operation status (such as low activity, medium activity, or high activity). It can be implemented through clustering or segmented thresholding, thereby achieving a structured characterization and dynamic identification of the port's economic operation status.

IV. RESULTS AND DISCUSSION

A. EXPERIMENTAL DATA AND EXPERIMENTAL ENVIRONMENT CONFIGURATION

Selected experimental data included multi-source spatiotemporal data collected within a typical border port zone, namely the data of mobile signals, data on customs clearance procedures, data on traffic checkpoint flows, and data on consumer activities. These spatiotemporal data span several consecutive time periods and the key functional area of the port zone along with its adjacent radiation zone.

As part of the data preprocessing, temporal normalization and spatial rasterization were performed, after which the selected data were consistently represented in a structured spatiotemporal data input format. For performing experiments, a powerful computational resource was used, equipped with multi-core processors and graphics processing units in order to ensure the operation of the spatiotemporal graph neural network.

B. COMPARISON MODEL SETTINGS

In order to test the performance of the proposed model, several typical approaches have been employed for benchmarking based on three types of methods: statistical approach,

traditional machine learning approach, and deep learning approach.
(1) Statistical approach: Historical mean model and moving average model are employed for describing the trends of the original time series.
(2) Machine learning approach: Random Forest model and support vector regression are employed for performing non-linear fitting for flow properties.
(3) Deep learning approach: LSTM and Graph Convolution Network models are employed for modeling the temporal and spatial dependencies.

All the benchmarking models were tested on the same dataset under the same conditions. Crucially, to ensure a fair evaluation of the architectural advantages of ST-GNN, all baseline models (Random Forest, SVM, LSTM, etc.) were provided with the same high-quality inputs, including the fused multi-source spatiotemporal fea tures and multi-scale data components described in Section 3.1. This ensures that the observed performance gains of the proposed method stem from its superior ability to model complex spatiotemporal dependencies rather than differences in data preprocessing.

C. EVALUATION INDEX SYSTEM

To comprehensively evaluate the model's performance in predicting population movement and estimating economic activity, a multidimensional evaluation index system is constructed. For predictive performance, mean squared error (MSE) and mean absolute error (MAE) are used for evalu ation, and their calculations are as follows:

$$\mathrm{MSE} = \frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2 \tag{19}$$

$$\mathrm{MAE} = \frac{1}{n}\sum_{i=1}^{n}|y_i - \hat{y}_i| \tag{20}$$

Where (y_i) represents the actual value and $(\hat{y}_i)$ represents the predicted value.

Regarding the consistency of economic activity estimates, the Pearson Correlation Coefficient is introduced to measure $r = \frac{\sum_i (E_i - \overline{E})(\hat{E}_i - \hat{E})}{\sqrt{\sum_i (E_i - \overline{E})^2}\sqrt{\sum_i (\hat{E}_i - \overline{\hat{E}})^2}}$ the degree of consistency between predicted economic activity and actual economic performance.

D. EXPERIMENTAL VARIABLES AND ABLATION SETTINGS

To investigate the contribution of each individual component in the proposed model, an ablation study with hierarchical removals of core components is adopted.

Firstly, the spatiotemporal graph structure is removed while the other two parts remain unchanged to examine the effect of modeling spatial dependency. Secondly, the multi-scale modeling part is removed and the prediction process is based on only one scale to assess the effect of time decomposition.

Finally, the causal lag analysis part is removed and modeling is conducted by synchronous relations only to verify the effect of the time delay mechanism.

Through comparison of these groups of models, the marginal contribution of each component will be eval uated.

From Figure 1, one can see that there are substantial discrepancies in the performance of various models when dealing with the issue of forecasting the population movement. On the whole, in comparison with traditional statistical approaches, deep learning techniques, as well as models based on spatiotemporal data analysis, demonstrate a positive trend toward improved prediction accuracy. Specifically, the historical mean model and moving average model, used as benchmarking strategies, exhibit relatively high mean squared errors (MSE) and mean absolute errors (MAE): 0.218 and 0.193; 0.341 and 0.312, respectively, which indicates that purely historical statistical data cannot be used to describe the complicated nature of changes in the population movement at ports. In turn, the spatiotemporal graph neural network model suggested in this work performs excellently when considering the factors of time and space simultaneously, with $\mathrm{MSE} = 0.071$, $\mathrm{MAE} = 0.174$, and $\mathrm{RMSE} = 0.266$, far outperforming all other comparative models in terms of the prediction accuracy and demonstrating the potential of the approach to effectively describe the characteristics of population flow. Even though the execution time of the proposed model is higher (4.38 s), its computational complexity is still reasonable compared to performance improvement.

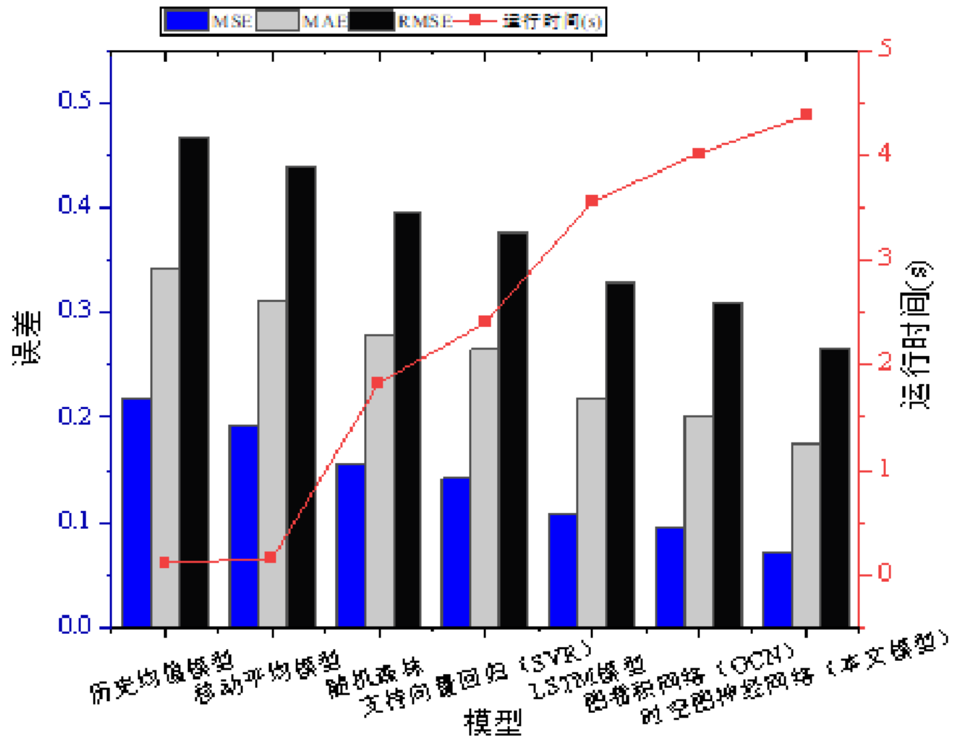

FIGURE 1. Comparison of the Prediction Performance of Different Models for Population Movement

TABLE 1. Results of Economic Vitality Assessment and Ablation Experiment

Model variants	Economic vitality correlation coefficient r	MAE (Economic vitality)	MSE (Economic vitality)	Illustrate
Basic Time Series Model	0.782	0.214	0.089	Time-only modeling
+Spatial modeling (structure without graph)	0.836	0.187	0.072	Add spatial features
+Graph structure modeling	0.891	0.154	0.058	Introducing a mobile network
Multi-scale modeling	0.918	0.132	0.046	Add short-term/long-term decomposition
+Causal lag analysis (complete model in this paper)	0.942	0.118	0.039	Complete method

As can be seen from Table 1, there is an evident trend of gradual optimization in the performance of various model variants on the economic vitality evaluation task, which means that the contribution of each module to the performance of the model is quite significant. For instance, the basic time series model depends solely on the temporal dimension of information with the value of $r = 0.782$, $\text{MAE} = 0.214$, and $\text{MSE} = 0.089$. These numbers indicate that it is impossible to characterize the driving mechanism of economic vitality using the basic time series model alone since its overall performance is not high enough. Therefore, implementing the multi-scale modeling technique further improves the performance of the model with the correlation coeffi cient of 0.918, MAE of 0.132, and MSE of 0.046. It can be seen that such modeling increases the adaptability of the model to various complex time dynamics by decomposing short-term fluctuations into long-term trends. Finally, after implementing the causal lag analysis module, the full-fledged model implemented in this study provides the highest level of performance with the correlation coefficient of 0.942, MAE of 0.118, and MSE of 0.039.

V. CONCLUSION

This work aims at investigating the influence of port population flow on the economic vitality of a region. Based on spatiotemporal big data, we have developed a robust analytical framework that synthesizes popu lation dynamics and regional economic indicators through the lens of spatiotemporal big data analysis. In addition, multi-source and spatiotemporal alignment techniques are adopted to establish a structured port population flow and a spatiotemporal flow network used for depicting the migration relationship between regions. The spatiotemporal graph neural network and multiscale dynamic modeling are employed to model and predict the process of population flow. Besides, a method for assessing economic vitality in view of multi-dimensional indexes including flow intensity, dwell time, and frequency of crossing border is developed. What is worth mentioning is that the coupling relationship and the path of influence from population flow change to economic vitality change with respect to different lags are explored. The experimental results indicate the proposed approach has better prediction and assessment capabilities for population flow and economic vitality. Nevertheless, there are still some deficiencies in this work in modeling complicated factors, such as unexpected events or policy changes.

AUTHOR CONTRIBUTIONS

Conceptualization – Ximeng Zhang and Yuan Yao; methodology – Ximeng Zhang and Yuan Yao; investigation – Ximeng Zhang and Yuan Yao; writing-original draft preparation – Ximeng Zhang and Yuan Yao. All authors have read and agreed to the published version of the manuscript.

CONFLICTS OF INTEREST

The authors declare no conflict of interest.

FUNDING

This research was supported by, Project:Research on Marine Economic Development and Transformation and Upgrading in Hunchun JJKH20261034SK

ACKNOWLEDGEMENTS

Not applicable.

REFERENCES

[1] K. Tao, X. Zhao, Y. Lan, et al., "Spatiotemporal pattern analysis of population movement during the Spring Festival travel rush in China based on Tencent migration data," Integrated Technology, vol. 12, no. 1, pp. 68–78, 2023.

[2] H. Jiang, W. Fang, T. Xu, et al., "Indoor evacuation optimal path planning model based on Dijkstra's algorithm," Journal of Tsinghua University (Natural Science Edition), vol. 65, no. 4, pp. 742–749, 2025.

[3] H. Liu, W. Bi, C. Wu, et al., "Spatial distribution characteristics and influencing factors of enterprise algorithm collusion index in the intelligent era," Economic Geography, vol. 45, no. 3, pp. 44–51, 2025.

[4] G. De, "Application of Artificial Intelligence in Human Resource Data Analysis," Modern Economic Management, vol. 6, no. 3, pp. 92–94, 2025.

[5] X. Wang, Q. Dong, X. Li, et al., "A method for passenger flow detection and personnel search in rail transit based on an improved Grounding DINO model," Railway Computer Applications, vol. 34, no. 11, pp. 61–68, 2025.

[6] Y. Wang, H. Luo, and G. Yang, "Analysis and Evolution of Inter-provincial Mobile Network of Scientific Researchers," Scientific Research Management, vol. 43, no. 3, pp. 79–88, 2022.