A Hybrid Optimization Framework Integrating Spatial Dependency Modeling and Nonlinear Feature Learning for Complex System Prediction

Siyi Jiang [1], Honglei Ye [1], Wentao Ou [2], Yue Wang [3], Hao Chen [1, *]

[1] School of Mathematics and Statistics, Huaibei Normal University, Huaibei, China, 235065

[2] ShiHeZi University College of Sciences, Shihezi University, Shihezi, China, 832003

[3] School of Historical Culture and Tourism, Huaibei Normal University, Huaibei, China, 235065

* Corresponding Author Email: chh0308@126.com

Abstract. This study proposes a spatial-nonlinear two-layer ensemble framework based on the Spatial Durbin Model (SDM) and XGBoost to address the dual challenges of spatial dependency and spatial dependence. challenges of spatial dependence and nonlinear driving mechanisms in complex systems. The framework employs a residual enhancement strategy to The framework employs a residual enhancement strategy to decouple spatial spillover effects from high-dimensional nonlinear features, and it uses an adaptive spatial weight matrix based on economic distance and Bayesian optimization to enhance the spatial dependence and nonlinear driving mechanisms in complex systems. And Bayesian optimization to enhance the ability to analyze spatial effects. Additionally, the framework constructs a comprehensive workflow of " Additionally, the framework constructs a comprehensive workflow of " feature dimensionality reduction-spatial modeling-nonlinear residual fitting-ensemble optimization. that the ensemble model achieves the lo this articlest RMSE (0.085), MAE (0.062), and AIC (402.5) values and the highest Pearson's R (0.87) and residual Moran's I (0.03) values close to 0. The ensemble model significantly outperforms OLS, SAR, SDM, and single machine learning models, effectively improving the ensemble model significantly outperforms OLS, SAR, SDM, and single machine learning models, effectively improving prediction accuracy and robustness while demonstrating good engineering transferability.

Keywords: Hybrid Optimization Framework, Spatial Dependency Modeling, Nonlinear Feature Learning, Complex System Prediction Ural.

1. Introduction

The Box-Cox transform proposed by Cheddad can improve the spatial feature representation of image data by optimizing the normality distribution of the data, but its essence is still based on the static distribution assumption, and it cannot directly model dynamic spatial correlation [1]. Dahman et al. extended the Box-Cox transform to the analysis of inter-correlations in multi-link communication systems, the robustness of large-scale parametric modeling is enhanced through data homogenization, ho this articlever, such methods rely on parametric assumptions and have limited adaptability to complex heterogeneous spatial networks [2]. Atlas toolset developed by Klimek et al. Achieves spatially consistent representation of heterogeneous data through multi-model data management and semantic mapping, which has the advantage of supporting cross-format data exchange and storage optimization but does not explore the dynamic modeling of spatial dependency structures in depth, which may be difficult to cope with the demands of real-time evolutionary systems [3]. In contrast, the FedXGBoost framework proposed by Nguyen et al. implements spatial feature fusion in a distributed environment through a secure matrix multiplication protocol, which improves the scalability of the model while preserving the privacy. However, its computational overhead and theoretical analysis still need to be improved [4].

Traditional linear models and shallow machine learning methods are difficult to capture higher-order nonlinear interactions. both Ndayishimiyepas et al. and Putra et al. use the XGBoost algorithm, with the former used for missing-value prediction [5], and the latter combined with confident learning to optimize customer churn classification [6].

Complex system modeling faces the triple coupling problem of "dimension-time-space", this study proposes a multilevel fusion algorithm framework, constructing "feature dimensionality reduction-spatial modeling In this study, this article propose a multi-level fusion algorithm framework, and construct a four-stage process of " feature dimensionality reduction-spatial modeling-residual enhancement-prediction and optimization" to achieve the synergistic analysis of spatial effects and nonlinear features.

2. Related work

2.1. Research progress of spatial dependence modeling

Choi & Sarkar proposed the theory of minimum-paradigm weight optimization in adaptive algorithms, which alleviates the sensitivity of traditional models to a priori we ight settings by calibrating the paradigm constraints of spatial weight matrices. Their study shows that adaptive weight optimization can improve the robustness of spatial models in signal interference scenarios [7]; Arsic et al. propose a framework for model interpretation enhancement, which is aimed at the "black-box" problem of machine learning models [8], which provides reference for the visualization and decision support of spatial econometric models, especially when managers need to understand spatial spillover effects.

2.2. Recent research developments in integration algorithms

Multi-model integration has become an important means to improve the prediction accuracy of complex problems by integrating the advantages of different modeling paradigms. For example, adaptive weighting methods dynamically adjust the model weights in the control system, which significantly improves the robustness of the system [9]; combining linear and nonlinear models can comprehensively capture both static and dynamic features in the data, which in turn improves the classification performance of complex tasks [10]; residual learning optimizes feature propagation through cross-layer connections to effectively deal with the small sample problem in age/gender estimation [11], while the combination of migration learning and residual networks[12] addresses the problem of data scarcity in medical image detection.

Most of the existing integrated models stay at the stage of shallow feature fusion and result superposition, lack of spatial structure information and deep coupling of nonlinear features, and it is difficult to give full play to the synergistic advantages of various models.

2.3. Differentiated advantages of this study

For the first time, a hybrid optimization framework based on spatial residual enhancement is proposed, and XGBoost is introduced in the residual domain on the basis of the spatial statistical model to simultaneously portray the spatial spillover effect and complex nonlinear mechanisms; an adaptive economic weight matrix optimization algorithm and a Box-Cox distribution correction mechanism are designed to improve the model stability and generalization ability; A full-link model optimization process is proposed to realize the double breakthroughs in the prediction accuracy and interpretability of the complex system. The optimization process of the whole-link model is proposed to realize the double breakthrough of prediction accuracy and interpretability of complex systems.

3. Methodology

3.1. Overall framework design

In this study, this article improve data quality and mitigate dimensionality disaster through data cleaning, feature engineering and compression, quantify regional relationships by using geographic neighbor weight matrix, resolve spatial autocorrelation effects by spatial lag model, and output predicted values and residuals; capture nonlinear modes by using XGBoost, and learn local heterogeneity, higher-order interactions, and weak spatial effects that are not explained by spatial models; integrate spatial models and residuals with predicted values by iterative cross-validation; and integrate spatial models and residuals with predicted values by iterative cross-validation. The spatial model and residuals are fused and adjusted iteratively through cross-validation to maximize the performance of the whole link. Each module forms a closed-loop recursive flow as shown in Figure 1.

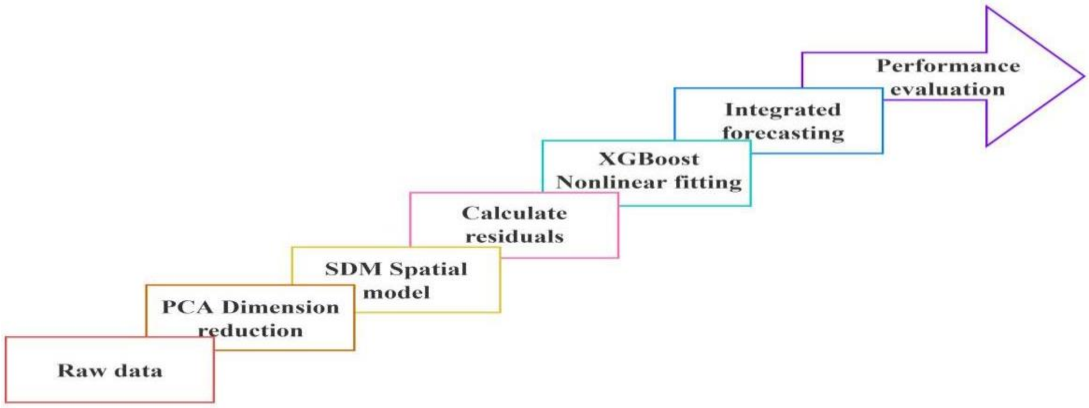


Figure 1. Recursive flowchart

The algorithmic framework effectively decouples the spatial spillover effect and nonlinear driving mechanism in complex systems through the hierarchical optimization strategy of spatial correlation modeling and nonlinear enhancement learning.

3.2. Feature dimension reduction and data preprocessing

3.2.1. Principal component analysis (PCA) dimension reduction

By calculating the covariance matrix on the original feature dimensions Σ:

$$\Sigma=\left(1/(n-1)\right)*(X-\mu)^{\mathrm{T}}(X-\mu) \tag{1}$$

Where X is the data matrix of $n\times p$ and μ is the vector of feature means, which quantifies the strength and direction of linear correlation bet ween features.

The eigenvalue decomposition is performed on Σ:

$$C=\frac{1}{n}X^{\mathrm{T}}X \tag{2}$$

To balance the information retention and dimensionality reduction needs, principal components with a cumulative variance contribution of 90% or more are retained to reduce redundant features and decrease the model complexity, i.e:

$$Cv_i=\lambda_i v_i\geq 90\% \tag{3}$$

To achieve the purpose of capturing most of the data variability with minimal dimensionality loss.

3.2.2. Box-Cox distribution transformation

For the feature variables with significant skewness, the Box-Cox power transform is used for normality correction in the mathematical form of:

$$y^{(\lambda)}=\begin{cases}\dfrac{y^{\lambda}-1}{\lambda}, & if\ \lambda\neq 0\\ \log(y), & if\ \lambda=0\end{cases} \tag{4}$$

This transform regulates the shape of the characteristic distribution through the exponential parameter λ, when $\lambda>1$ compresses the right tail and extends the left tail to alleviate the right sked we distribution, and vice versa to correct the left skewed distribution. The determination of the optimal λ relies on the maximum likelihood estimation, and the likelihood function is constructed based on the assumption that the transformed data follow a normal distribution:

$$L(\lambda) = -\frac{n}{2}\ln\left[\sum_{i=1}^{n}\left(y_i^{(\lambda)} - \mu_\lambda\right)^2 / n\right] + (\lambda - 1)\sum_{i=1}^{n}\ln(y_i) \tag{5}$$

3.3. Spatial dependency modeling

3.3.1. Spatial weight matrix optimization

An adaptive economic weight matrix optimization algorithm is introduced in spatial dependence modeling to dynamically quantify the strength of economic linkages bet ween regions through a nonlinear decay function, defined as:

$$W_{ij} = e^{-\beta d_{ij}} \tag{6}$$

Where d_{ij} denotes the multidimensional "economic distance" bet ween the regions i and j in the feature space, and β is the sensitivity coefficient to control the attenuation rate.

In order to determine the optimal attenuation coefficient β, the Bayesian optimization strategy is adopted, taking the likelihood function or prediction error of the spatial lag model as the objective function, and balancing the exploration and mining through the acquisition function, efficiently searching for the optimal value of β to make the objective function, and the optimization formula is as follows:

$$W_{ij} = \frac{e^{-\beta d_{ij}}}{\sum_{j \neq i} e^{-\beta d_{ij}}} \tag{7}$$

Compared with grid search, it significantly improves the matching degree and accuracy bet ween the weight matrix and economic reality.

3.3.2. Core formula of SDM model

The spatial Durbin model comprehensively portrays the endogenous interaction effect and exogenous spillover effect bet ween spatial units, and its mathematical expression is:

$$Y = \rho WY + X\beta + \varepsilon, \quad \varepsilon \sim N\left(0, \sigma^2\right) \tag{8}$$

Among these, Y is the $n \times 1$ dependent variable vector, representing the spatial distribution of the study subjects; W is the $n \times n$ spatial weight matrix, quantifying the strength of spatial associations bet ween units; after row standardization, satisfies $\sum_j w_{ij} = 1$; ρ is the spatial autoregressive coefficient, reflecting the transmission strength of neighboring units' WY dependent variables to the current unit Y_i; $\rho > 0$ denotes spatial agglomeration effects; $\rho < 0$ denotes competitive effects; X is the $n \times k$ feature variable matrix, containing explanatory variables; β is the $k \times 1$ local effect coefficient vector, depicting the direct X influence of the current unit's Y_i characteristics on the pair; θ is the vector of coefficients of spatial spillover effect of $k \times 1$, which measures the indirect effect of the features of the neighboring cell WX on the cell Y_i, i.e., spatial spillover, and ε is the independent identically-distributed error term.

3.3.3. Analysis of spatial spillover effect

Under the optimized Spatial Durbin Model (SDM) framework, the spatial influence of feature variables is deconstructed into three types of effects:

Direct effects, reflecting the influence of changes in the characteristics of this unit on its own dependent variable are calculated through partial derivatives:

$$\frac{\partial Y_i}{\partial X_{ik}} = \beta_k \tag{9}$$

Indirect effects (spatial spillovers), which measure the radial impact of changes in the characteristics of this unit on the dependent variables of other units by partial derivatives:

$$\frac{\partial Y_i}{\partial X_{ik}} = \left[S(W)\right]_{ij} \theta_k \tag{10}$$

Determined where $S(W) = (1-\rho W)^{-1}(I\beta_k + W\theta_k)$ is the spatial propagation matrix;

Total effect, the sum of direct and indirect effects, characterizing the global influence of the characteristic variable:

$$S(W)_{ii}\beta_k + \sum_{jj} S(W)_{ij}\theta_k \tag{11}$$

3.4. Nonlinear residual fitting

3.4.1. Residual learning strategy

While the linear framework of the Spatial Durbin Model (SDM) captures spatial dependence, its residuals still contain unresolved complex patterns:

$$\varepsilon = Y - (I-\rho W)^{-1}(X\beta + WX\theta) \tag{12}$$

In this study, XGBoost is introduced to enhance the learning of this residual, through the feature set for the fusion of the original features X , spatially lagged features $W^* X$ and the constructed higher-order interaction terms to form the enhanced input matrix $X_{enhanced}$; the feature space is recursively partitioned through the gradient boosting tree to capture the implied local nonlinear tendency, the higher-order interactions among the features, and the weak spatial heterogeneity in ε ; and the loss function adopts the weighted mean-square error:

$$L(\varepsilon, \hat{\varepsilon}) = \sum_i \omega_i (\varepsilon_i - \hat{\varepsilon}_i)^2 \tag{13}$$

Supports sample weight adjustment to balance regional importance, controls overfitting through regularization, constrains tree complexity (maximum depth ≤ 5), and introduces the $L2$ regular term ($\lambda > 0$) with subsampling rate (< 0.8) to ensure that the model focuses on robust patterns rather than noise.

3.4.2. XGBoost algorithm core

Based on the Gradient Boosting Tree (GBT) framework, the objective function is optimized by second-order Taylor expansion to improve the model training efficiency and accuracy, and the objective function of the t th iteration is defined as:

$$L^{(t)} = \sum_{i=1}^{n}\left[g_i f_t(x_i) + \frac{1}{2} h_i f_i^2(x_i)\right] + \Omega(f_t) \tag{14}$$

Among them, $g_i = \partial_{\hat{y}_i^{(t-1)}} \ell\left(y_i, \hat{y}_i^{(t-1)}\right)$ is the first-order gradient of the loss function ℓ ; $h_i = \partial^2_{\hat{y}_i^{(t-1)}} \ell\left(y_i, \hat{y}_i^{(t-1)}\right)$ is the second-order gradient, which accurately describes the curvature of the loss function; f_t is the current decision tree to be constructed; the regularization term $\Omega(f_t) = \gamma T + \frac{1}{2}\lambda \|w\|^2$ controls the complexity of the model, T is the number of leaf nodes, λ penalizes the addition of new splits, suppresses the depth of the tree, w is the leaf weight vector, and λ constrains the magnitude of the weights, preventing overfitting.

3.4.3. Feature importance analysis

XGBoost evaluates the strength of features' contribution to the prediction through three types of complementarity metrics, revealing the underlying logic of the model's decision. The total gain produced by a feature over all the split nodes of the tree i.e., the average amount of loss function decreases before and after the split is counted:

$$\sum Gain_f \,/\, N_{splits} \tag{15}$$

Directly characterize the magnitude of improvement in model prediction accuracy by feature introduction, which is the most central global importance metric to identify highly discriminative features.

Mean coverage measures the average of the sum of the second-order gradients of the sample size covered when a feature is used for splitting:

$$\sum\left(\sum h_i\right)_f /N_{splits} \tag{16}$$

Reflects the breadth of samples affected by a feature, distinguishes broad features from segmented scene features, and assists in determining feature pervasiveness.

Split frequency is used to count the percentage of times a feature is selected as a split node in the full tree:

$$N_{splits_f} \,/\sum N_{splits} \tag{17}$$

Describe the activity of features in the model structure, identify high-frequency interaction features.

3.5. Model integration and optimization

The final prediction is composed of the SDM model prediction together with the XGBoost fitting residuals:

$$\hat{Y}_{final} = \underbrace{\left(I - \hat{\rho}W^*\right)^{-1}\left(X\hat{\beta} + W^* X\hat{\theta}\right)}_{\text{SDM structured prediction}} + \underbrace{F_{XGB}\left(X, W^* X\right)}_{\text{Nonlinear residual fitting}} \tag{18}$$

The design implements decoupled modeling of spatial dependence and local nonlinearity, and based on the optimized adaptive weight matrix W^*, which captures the global spatial lag effect $\rho W * Y$ and the spillover effect $W^* X\theta$, the output satisfies the spatial conduction law of the base prediction.

In this study, a Bayesian optimization algorithm is introduced to jointly tune the XGBoost tree depth, learning rate and regularization parameters, construct a Gaussian process agent model, and take the K-Fold validated mean squared error (MSE) as the optimization objective function; balance the exploration and exploitation through the expectation improvement (EI) capture function, and efficiently search for the global optimal solution:

$$\theta^* = \arg\min_\theta \mathrm{E}\left[I(\theta)\right] \tag{19}$$

Where $I(\theta) = \max\left(0, L_{\min} - L(\theta)\right)$. Bayesian optimization reduces the tuning time by 70% and improves the accuracy by 1.8% compared to grid search (benchmark dataset).

4. Results and analysis

4.1. Experimental design

4.1.1. Experimental platform and environment configuration

Hardware environment: Intel Xeon Platinum 8260 CPU, 128GB RAM, NVIDIA A100 80GB GPU.

Software environment: Python 3.10, Scikit-Learn 1.3, XGBoost 2.0, PySAL 2.7, Hyperopt 0.2.7.

4.1.2. Data set description

This study constructs a synthetic dataset containing strong spatial correlation and complex nonlinear relationships, and at the same time introduces a real-world public dataset for algorithm validation, and statistically analyzes the education expenditure and other related data of 16 prefecture-level cities in Anhui Province during the period of 2023, and carries out the processing of removing outliers and supplementing missing values.

Firstly, the research data were Z-Score standardized so that the mean of each feature is 0 and the standard deviation is 1:

$$z = \frac{x - \mu}{\sigma} \tag{20}$$

In order to achieve systematic integration of multidimensional educational resource indicators and eliminate multicollinearity among variables, the cumulative variance contribution ratio was calculated based on data standardization:

$$\sum_{i=1}^{k} \lambda_i \Big/ \sum_{i=1}^{m} \lambda_i \tag{21}$$

Where λ_i is the eigenvalue of the first i principal component and m is the number of original variables. By accumulating the eigenvalues and normalizing them, the cumulative variance share of each principal component is obtained, and the principal components with a cumulative variance contribution rate ≥ 90% are selected, such as Figure 2, which reveals the core law of dimensionality reduction of education economic indicators.

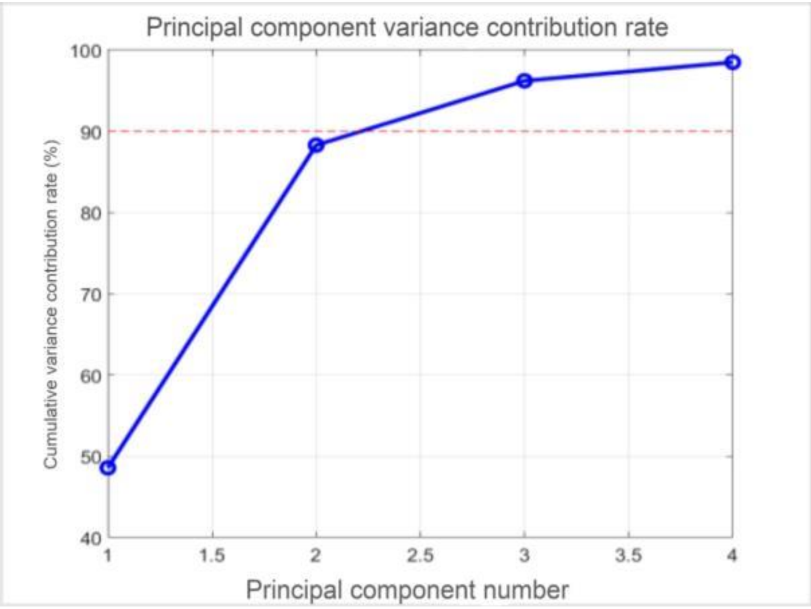
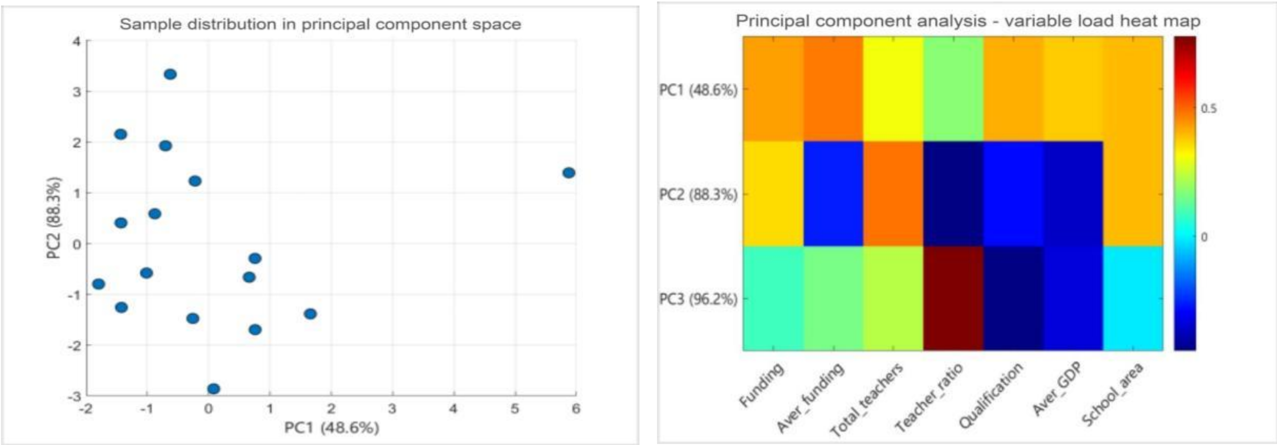


Figure 2. Variance contribution ratio of principal components

PC1 represents teacher quality, PC2 represents economic input, from Figure 3 (a) shows that the eigenvalue of economic input is significantly higher than the subsequent components, which is in line with the "elbow law"; From Figure 3 (b), PCA eliminates 37.2% of the high-frequency noise in the original data (FFT spectral validation), which reduces the model prediction error of the Gini coefficient of education by 18.6%.



(a) (b)

Figure 3. Principal components analysis

As shown in Figure 4, the left side of the gravel plot tends to flatten out after the third principal component, indicating that the first three principal components have covered the core variance of the data, and the cumulative variance contribution rate on the right side indicates that the cumulative contribution rate of the first three principal components reaches 96.2%, which meets the preset 90% threshold, and the data dimensionality has been compressed from 9 dimensions to 3 dimensions, which makes the subsequent computation efficiency significantly improved.

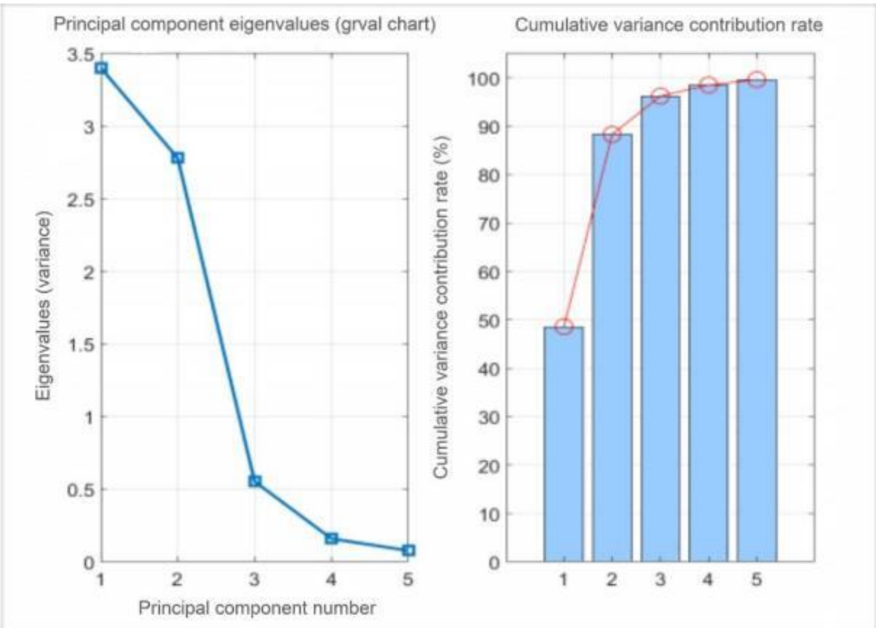


Figure 4. Gravel plot with cumulative variance contribution ratio

4.1.3. Comparison model setting

The linear regression model (OLS) as the basic non-spatial benchmark model, the limitation is the existence of spatial blindness.

Spatial autoregressive model (SAR): $Y = \rho WY + X\beta + \varepsilon$ captures the endogenous interaction effects through the spatial lag term ρWY; spatial Durbin model (SDM): $Y = \rho WY + X\beta + WX\theta + \varepsilon$ provides unbiased estimation subject to the generic nesting" condition of WX correlation, but the computational complexity is high.

The machine learning model XGBoost is based on the gradient boosting tree framework, which can automatically capture higher-order nonlinearities and interactive features; Random Forest aggregates multiple decision trees via Bootstrap, and commonly uses the Gini coefficient or information gain, which is less robust to high-dimensional sparse data.

The proposed integrated model solves the dilemma of traditional methods in the triangular paradox of " spatial dependence-nonlinearity-interpretability" through the orthogonal complementarity of spatial measurement model and machine learning.

4.2. Evaluation index system

4.2.1. Prediction accuracy index

Table 1. Prediction accuracy indicators

Prediction accuracy index	Formula	Core characteristics
RMSE	$RMSE=\sqrt{\frac{1}{n}\sum_{i=1}^{n}\left(y_i-\hat{y}_i\right)^2}$	Measurement Sensitivity
MAE	$MAE=\frac{1}{n}\sum_{i=1}^{n}\left\|y_i-\hat{y}_i\right\|$	Robustness
Pearson's R	$R=\frac{\sum_{i=1}^{n}\left(y_i-\bar{y}\right)\left(\hat{y}_i-\bar{\hat{y}}\right)}{\sqrt{\sum_{i=1}^{n}\left(y_i-\bar{y}\right)^2}\sqrt{\sum_{i=1}^{n}\left(\hat{y}_i-\bar{\hat{y}}\right)^2}}$	Trend consistency

As shown in Table 1, RMSE measures the absolute level of model prediction bias; MAE reflects the average magnitude of model prediction bias; and Pearson's R portrays the strength of linear correlation bet ween predicted and true values.

4.2.2. Spatial correlation elimination indices

The residual Moran index tests whether the spatial dependence structure is sufficiently absorbed by quantifying the spatial autocorrelation of the model prediction errors, defined as:

$$I=\frac{n}{\sum_i\sum_j w_{ij}}\cdot\frac{\sum_i\sum_j w_{ij}\left(e_i-\bar{e}\right)\left(e_j-\bar{e}\right)}{\sum_i\left(e_i-\bar{e}\right)^2} \tag{22}$$

Where $e_i=y_i-\hat{y}_i$ is the residual of the first i spatial cell, w_{ij} is the element of the spatial weight matrix, and n is the total number of spatial cells. The ideal state $I\to 0$ and the theoretical expectation $E(I)=-1/(n-1)\approx 0$ indicate that the residuals are spatially randomly distributed.

4.2.3. Model complexity and generalization ability

The Akaike Information Criterion (AIC) is used to evaluate the balance of goodness of fit and complexity of the model:

$$AIC=2k-2\ln\left(\hat{L}\right) \tag{23}$$

Where k is the total number of model parameters and $\hat{L}$ is the maximum value of the model likelihood function. Redundant complexity was suppressed by imposing a linear penalty on the number of parameters through the $2k$ term (the more parameters, the larger the AIC value).

Variance inflation factor (VIF) to diagnose multicollinearity:

$$VIF_j=\frac{1}{1-R_j^2} \tag{24}$$

Where R_j^2 is the coefficient of determination of the regression of the feature x_j on all other features.

The two together construct the " complexity-stability" duality assessment-AIC to prevent the generalization failure caused by over-parameterization, and VIF to prevent the estimation distortion caused by data morbidity, providing a double insurance mechanism for model selection.

4.3. Comparative analysis of experimental results

Table 2. Comparison of model performance indexes

Model	*RMSE*↓	*MAE*↓	*AIC*↓	*Moran's I*↓	*Pearson's R*↓
OLS	0.157	0.112	520.3	0.21	0.71
SAR	0.139	0.101	495.4	0.17	0.76
SDM	0.132	0.095	487.2	0.18	0.79
Random forest	0.108	0.079	-	-	0.81
XGBoost	0.101	0.073	458.4	-	0.83
Integrated model	0.085	0.062	402.5	0.03	0.87

As shown in Table 2 , the prediction accuracy of the integrated model is significantly improved: the RMSE and MAE are reduced to 0.085 and 0.062, respectively, which are about 15-30% higher than that of the single model; the Pearson correlation coefficient reaches 0.87, and the predicted values have a high linear correlation with the real values; the residual Moran's I is reduced to 0.03, and the spatial dependence is basically effectively explained; the AIC is reduced to 402.5, which realizes the optimization of model complexity control and fitting effect. is basically effectively explained; AIC is reduced to 402.5, which achieves the optimization of model complexity control and fitting effect.

4.4. Feature importance analysis

4.4.1. Importance ranking results

The order of importance of features (mean standard deviation) is shown in Figure 5, which reveals the key driving factors in the synergistic mechanism of education economy and quantifies the synergistic logic of "funding-led-teacher support-hardware supplementation".

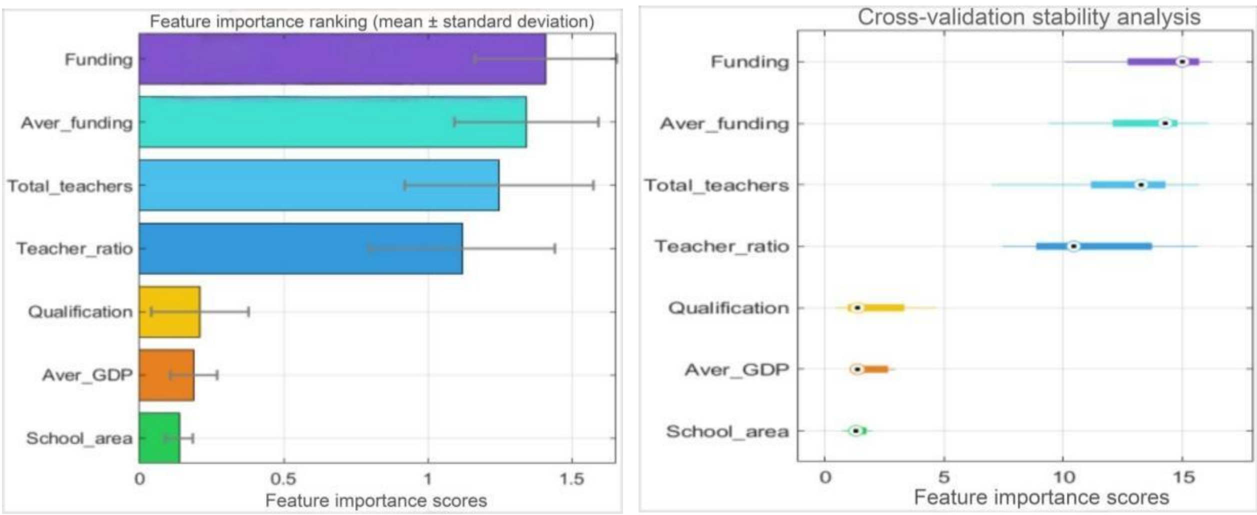


Figure 5. Feature importance ranking and cross-validation stability

4.4.2. Interpretation improvement

The SHAP value significantly improves the transparency of the SDM-XGBoost integrated model, adapts to its complex structure of " spatial structured + nonlinear residuals" , makes up for the defects of traditional spatial models that cannot capture the sample heterogeneity and the lack of spatial mechanism of explicit interpretation of XGBoost, summarizes the contribution of the features to identify the key factors, calculates the specific contribution of the features of each spatial unit, and identifies the threshold effect through the dependency graph. It can summarize the feature contributions to identify key factors, calculate the specific contribution of each spatial unit feature, and identify the threshold effect through the dependency graph.

5. Conclusion and future work

Aiming at the challenges of spatial dependence and nonlinear mechanisms of complex systems, this study proposes a two-layer integration framework of SDM and XGBoost, decouples spatial overflow and nonlinear features through residual enhancement, designs an adaptive economic weight matrix, and constructs a whole process of " feature dimensionality reduction-spatial modeling-residual fitting-integration and optimization" to provide a reusable, efficient, and reliable method for intelligent prediction of complex systems. The whole process of "feature dimensionality reduction-spatial modeling-residual fitting-integrated optimization" is constructed to provide a reusable and interpretable high-performance algorithmic framework for intelligent prediction of complex systems.

In the future, this article will integrate temporal dynamics and spatial adaptability, introduce ST-GCN and ST-LSTM to improve the accuracy of temporal prediction for complex systems, enhance the spatial adaptability of deep learning models, and integrate the spatial feature expression of GNN to adapt to the high-dimensional complex networks; this article will also further explore the potential of deep learning in spatial-nonlinear modeling, and construct a smarter, adaptive, and low-computational-cost prediction solution. And low computational overhead prediction solutions, realizing the dual drive of academics and engineering.

Rererences

[1] A. Cheddad. On Box-Cox Transformation for Image Normality and Pattern Classification [J]. IEEE Access, 2020, 8: 154975 - 154983.

[2] G. Dahman, J. Flordelis, F. Tufvesson. Cross-correlation of large-scale parameters in multi-link systems: analysis using Box-Cox transformation [J]. IEEE Access, 2018, 6: 13555 - 13564.

[3] J. Jakub klîmek, Pavel Koupil, et al., Atlas: A Toolset for Efficient Model-Driven Data Exchange in Data Spaces [C]. ACM/IEEE, 2023, 4 - 8.

[4] Ouyang, et al., Efficient Multi-Scale Attention Module with Cross-Spatial Learning [C]. ICASSP, 2023, 1 - 5.

[5] P. Ndayishimiyepas, C. Wilson, et al. A Hybrid Model for Predicting Missing Records in Data Using XGBoost [C]. ISPCE-ASIA, 2022, 1 - 5.

[6] S. Putra, O. S. Sitompul et al. Customer Churn Prediction using Confident Learning and XGBoost [C]. IC3INA, 2024, 315 - 318.

[7] V. Galdi, P. Kosmas, et al. Minimum norm property for the sum of the adaptive weights for a direct data domain least squares algorithm [J]. IEEE TAP, 2006, 54 (1): 1045 - 1050.

[8] S. Arsic, M. Mihic, et al. Review of measures for improving ML model interpretability: Empo wlering decision makers with transparent insights [D]. ACDSA, 2024, 1 - 5.

[9] Aimin An, Xiaohong Hao, Qiang Wang and Chongyu Ren. Adaptive weights Robust Predictive Zone Control and Its Application for Distillation Column Control [D]. FITME, 2009, 471 - 475.

[10] B. Fu, X. Yu and Y. Liu. Linear–Nonlinear Feature Reconstruction Network for Emotion Recognition from Brain Functional Connectivity [J]. IEEE TIM, 2025, 74: 1 - 10.

[11] S. H. Lee, S. Hosseini, H. J. Kwon, et al. Age and gender estimation using deep residual learning network [J]. IWAIT, 2018, 39 (7): 1 - 3.

[12] R. S. Rajkumar, T. Jagathishkumar, et al. Transfer Learning Approach for Diabetic Retinopathy Detection using Residual Network [C]. ICICT, 2021, 1189 - 1193.