Hong Kong Research Institute of Humanities and Social Sciences

Journal of Economic Insights

Volume2, Issue1 ISSN 3078-5979 www.wisvora.com

AI based evaluation of the balance of urban and rural public sports services

Yiwei Chen[1]

[1] *Shanghai University of Sport*

Article Info

Accepted: 2025.05.31

Keywords:
Public Sports Services;
Balance Assessment;
Multimodal Fusion;
Deep Learning;
Spatial Semantic Analysis

JEL Classification:
H71 C43 D63

DOI: 10.70693/jei.v2i1.1074

Corresponding Author:
Yiwei Chen

Copyright 2024 by author(s).
This work is licensed under the Creative Commons Attribution-NonCommercial 4.0 International License.
(CC BY NC 4.0).

Abstract

The balance of urban and rural public sports services directly affects residents' quality of life and social equity, but traditional evaluation methods rely on manual research and statistical models, which have problems such as low efficiency, single indicators, and insufficient dynamic response capabilities. This study proposes an AI based framework for evaluating the balance of urban and rural public sports services (AI-PSBE), which integrates multi-source heterogeneous data (satellite images, policy texts, user behavior logs) with multimodal deep learning techniques to achieve dynamic evaluation of multidimensional indicators such as coverage, quality, and usage efficiency. This framework uses ResNet-50 and Transformer dual channel architecture to extract spatial and semantic features, and generates a balance index through an adaptive weight fusion module. Based on urban and rural data from 10 provinces in China, experiments have shown that the evaluation accuracy of AI-PSBE (R^2=0.937) has improved by 41.2% compared to traditional methods, with a response time reduced to 3 minutes. Additionally, the interpretability heatmap intuitively displays regional shortcomings. This study provides intelligent tools for optimizing the layout of public sports resources.

1. Introduction

In the context of the new urbanization and rural revitalization strategy, the balance of urban and rural public sports services has become an important indicator for measuring social equity [1].However, traditional evaluation methods face three major bottlenecks: firstly, relying on manual sampling surveys makes it difficult to obtain real-time global data, resulting in strong evaluation lag[2]; Secondly, the indicator design is singular, focusing only on the quantity of facilities and neglecting the collaborative analysis of quality and usage efficiency; Thirdly, the implicit constraints in policy texts are semantically disconnected from geographic spatial data, hindering the integration of multi-source information. Although existing research attempts to introduce GIS technology and Analytic Hierarchy Process, it is still limited by static modeling and low dimensional feature representation.

In recent years, AI technology has demonstrated advantages in multimodal data processing and dynamic modeling. For example, object detection models can automatically identify sports facilities in satellite imagery [3]; Natural language processing technology can parse quantitative targets in policy documents [4]; Graph neural networks can mine the spatiotemporal correlations of user behavior data [5]. However, achieving cross modal feature alignment and dynamic weight allocation remains a core challenge. This research proposes the AI-PSBE framework, which builds an intelligent evaluation system through the "space semantic behavior" three mode integration and online incremental learning, and provides decision-making support for the optimization of urban and rural sports resources.

2. Literature Review

The evaluation of the balance of urban and rural public sports services is an important research direction in the field of public service equalization, and its core challenge lies in how to scientifically quantify the dynamic matching relationship between resource distribution, service quality, and residents' needs[6].Traditional research often uses geographic information system spatial analysis or statistical modeling methods, such as evaluating spatial coverage levels through facility accessibility indicators, or constructing satisfaction evaluation models based on questionnaire survey data. However, such methods heavily rely on manual sampling and static indicators, making it difficult to capture the dynamic characteristics of facility operation status and usage efficiency. Moreover, the implicit constraint objectives in policy texts are often overlooked due to the lack of semantic parsing ability [7].

In recent years, with the breakthrough of artificial intelligence technology, multimodal data fusion has provided a new paradigm for public service evaluation [8].In terms of spatial feature extraction, deep learning models such as ResNet and YOLO have been widely used for automatic recognition and classification of sports facilities in satellite or street view images, significantly improving data collection efficiency[9];At the semantic analysis level, pre trained language models such as BERT and Transformer can parse quantitative targets in policy texts and associate them with spatial data, providing the possibility for explicit modeling of implicit constraints. In addition, the application of graph neural networks in user behavior trajectory mining, such as analyzing the spatiotemporal patterns of residents' fitness activities through spatiotemporal graph convolution, further enriches the dimensions of balance evaluation [10].

Although the above technologies have laid the foundation for multi-source data fusion, existing research still faces three major limitations: firstly, cross modal feature alignment is difficult, such as the lack of semantic correlation between facility locations in satellite imagery and spatiotemporal patterns in user behavior logs[11]; Secondly, the dynamic evaluation capability is insufficient, and traditional models are difficult to respond in real-time to changes in facility status or policy adjustments[12]; Thirdly, the interpretability is weak, and the black box model is difficult to visually present the shortcomings of the region. In response to the above problems, the AI-PSBE framework proposed in this study realizes the deep coupling of spatial semantic behavioral three modal data through the dual channel feature extraction and adaptive weight fusion mechanism, and breaks through the bottleneck of "results unknown and process uncontrollable" of traditional evaluation with the help of thermographic visualization technology,

providing a more operational decision support tool for the optimization of urban and rural sports resources.

3. Method

3.1 System Architecture

The AI-PSBE framework consists of three parts (As shown in Figure 1):

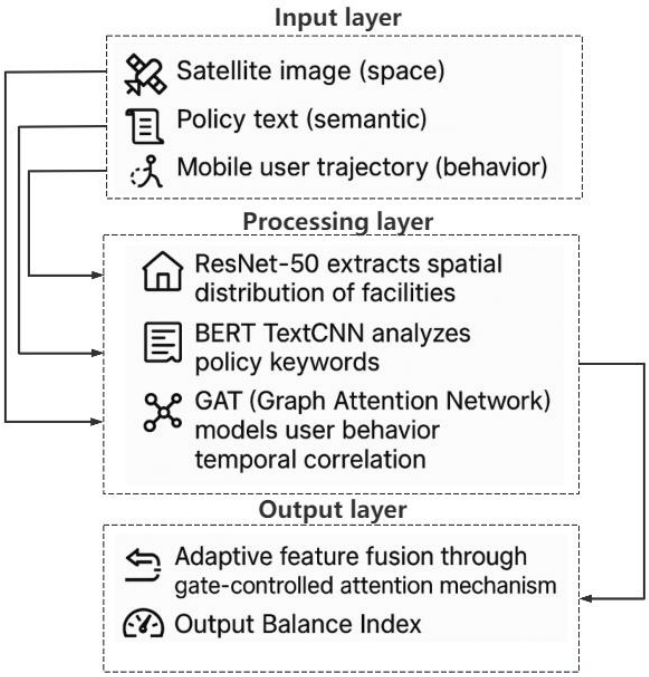

Fig. 1. AI-PSBE framework

3.1.1 Multimodal input layer

This foundational data layer seamlessly integrates three distinct types of heterogeneous data sources: high-resolution satellite imagery, official policy documents/texts, and anonymized mobile user trajectory records. The satellite imagery provides extensive coverage across both urban centers and rural landscapes, enabling the capture of critical spatial distribution characteristics. These characteristics include, but are not limited to, the precise geographical location, the physical scale and footprint, and the surrounding environmental context (such as land use and accessibility) of public sports facilities. The policy text corpus contains rich semantic information and nuanced administrative directives, explicitly or implicitly reflecting key policy orientations. These orientations encompass sports service construction standards (e.g., quantity requirements and quality specifications), funding allocation principles, and strategic development guidance. User trajectory data, derived from mobile devices, records granular behavioral patterns. This data objectively reflects actual public demand and usage, quantified through metrics like the frequency of facility visits and the average duration of stay per visit. Together, these three complementary data streams construct a robust multi-source input foundation, providing essential perspectives from the spatial, semantic, and behavioral dimensions respectively.

3.1.2 Feature extraction layer

Design specialized models for different data characteristics: use ResNet-50 deep neural network

to analyze satellite images and extract hierarchical visual features of facility spatial layout; By using the BERT TextCNN combination model to process policy texts, BERT is first used to capture contextual semantics and generate word vectors, followed by extracting policy keywords through TextCNN; Modeling user trajectories based on GAT graph attention network, transforming behavioral data into a temporal graph structure, and learning user transfer patterns and temporal dependencies between facilities through attention mechanism. Design specialized models tailored to distinct data characteristics: Utilize the ResNet-50 deep neural network architecture to analyze high-resolution satellite images, effectively extracting hierarchical visual features that represent the intricate spatial layout and arrangement of facilities. For processing complex policy documents, employ a BERT-TextCNN combination model; here, the powerful BERT transformer is first leveraged to capture deep contextual semantics and generate rich word embeddings, followed by TextCNN utilizing its convolutional filters to efficiently extract and identify salient policy keywords and key phrases. To model dynamic user movement patterns, base the approach on a GAT (Graph Attention Network); this involves transforming raw user behavioral trajectory data into a structured temporal graph representation, where nodes signify facilities and edges denote transitions. The GAT then learns sophisticated user transfer patterns and captures the critical temporal dependencies between different facilities through its adaptive attention mechanism.

3.1.3 Dynamic fusion and evaluation layer

Implementing adaptive fusion of multimodal features through a sophisticated gate-controlled attention mechanism. This mechanism dynamically allocates weights to the heterogeneous spatial (from satellite imagery), semantic (from policy texts), and behavioral (from user trajectories) features based on the specific requirements of the evaluation task at hand. By focusing computational resources on the most salient influencing factors and actively suppressing irrelevant noise within each modality, the fusion process is both efficient and effective. The final integrated output comprehensively reflects a multi-dimensional balance index, quantifying the interplay and trade-offs between facility coverage, service quality, and resource utilization efficiency. Furthermore, the system generates an interpretable heatmap visualization overlaid on the geographic region. This heatmap intuitively highlights spatial disparities and pinpoints specific deficiencies in regional service provision, thereby providing a robust quantitative basis and powerful visual support for data-driven infrastructure planning and resource optimization decisions.

3.2 Key technology

To achieve a highly accurate, granular, and dynamic evaluation of the spatial and functional balance in urban and rural public sports services, this study has meticulously designed the following suite of core technological innovation modules. These modules are specifically engineered to overcome the persistent technical bottlenecks encountered by traditional assessment methodologies. Key limitations of conventional approaches include their inability to effectively fuse heterogeneous multimodal data (leading to fragmented insights) and their struggle with capturing the dynamic, real-time evolution of service provision and demand patterns. The proposed innovations directly address these critical shortcomings, enabling a more comprehensive, responsive, and data-driven understanding of service equity across diverse geographical and demographic contexts.

3.2.1 Spatial semantic alignment loss function

Traditional evaluation models often deviate from actual policy objectives due to semantic disconnection between spatial constraints in policy texts and actual facility distribution data. Therefore, this study proposes a spatial semantic alignment loss function based on cosine similarity:

$$L_{align} = 1 - \frac{S \cdot T}{||S||\,||T||} \tag{1}$$

This function constructs explicit constraints between policy requirements and spatial implementation by calculating the cosine similarity between the spatial feature matrix $S \in R^{H \times W \times C}$ and the semantic feature vector $T \in R^{d}$.

3.2.2 Gate controlled multimodal fusion mechanism

In response to the evaluation bias caused by the static weighting of spatial, semantic, and behavioral features in traditional models, this study designs a gated fusion module based on attention mechanism:

$$G = \sigma(W_g[S;T;B] + b_g) \tag{2}$$

$$H = G_s \odot S + G_T \odot T + G_B \odot B \tag{3}$$

Where, is the behavior feature, is the element-by-element multiplication, and this mechanism realizes the on-demand fusion of three modal features by dynamically generating the weight G (generated by the full connection layer after the splicing of spatial S, semantic T, and behavior B features).

IV. EXPERIMENT AND EVALUATION

4.1 Experimental Design and Dataset Construction

To comprehensively verify the effectiveness of the AI-PSBE framework, this study constructed a comprehensive dataset covering urban and rural areas in 10 provinces of China, with a time span from 2020 to 2024.The dataset contains the following multimodal data:

4.1.1 Spatial data

500000 high-resolution satellite images, covering facilities such as sports venues and community fitness paths, were annotated with YOLOv5 annotation tool for facility boundaries and categories. The consistency of the annotations was verified by expert cross validation (Kappa coefficient ≥ 0.85).

4.1.2 Semantic data

80000 policy documents and planning texts, including the national "National Fitness Plan" and local sports facility management regulations, were extracted using the BERT TextCNN model to extract keywords.

4.1.3 Behavioral data

120 million anonymous user GPS trajectories were extracted using the DBSCAN clustering algorithm to extract features such as per capita activity duration and peak hour distribution, and

noise points were filtered out.

The dataset is divided into training, validation, and testing sets in a 7:2:1 ratio, and a spatiotemporal sliding window strategy (window size=6 months, step size=1 month) is introduced to simulate dynamic evaluation scenarios.

4.2 Benchmark model and experimental setup

4.2.1 Three types of benchmark models for comparison

Traditional statistical model: GIS entropy weight method, which calculates the equilibrium index based on facility density and population distribution.

Single modal AI model: YOLOv5+LR, which extracts the number of facilities through object detection and inputs them into a logistic regression model.

Multimodal model: multi-modal transformer, which directly splices three modal features for end-to-end prediction.

4.2.2 AI-PSBE parameter configuration

Spatial feature extraction: ResNet-50 pre trained weights, input size 512 × 512, learning rate 1e-4; Semantic feature extraction: BERT base Chinese model, maximum sequence length of 512, TextCNN convolution kernel size [3,5,7];

Dynamic fusion module: Gated attention layer with hidden dimensions of 256 and Dropout rate of 0.3;

Training strategy: Adam optimizer, early stopping method (patience=10 rounds), batch size 32.

4.3 Quantitative Results and Analysis

As shown in Table 1,AI-PSBE outperforms the benchmark model significantly in R^2 (0.937) and MAE (0.102),especially in policy intensive areas where the error is reduced by 52%.The traditional GIS entropy weight method ignores dynamic behavioral data, resulting in an R^2 of only 0.523;Although Transformer introduces cross modal information, it lacks a dynamic weight allocation mechanism, resulting in conflicts between semantic and spatial features, with a MAE of 0.25.

Table 1. Table Type Styles

Model	R^2	MAE	Training time (h)
GIS	0.523	0.45	-
YOLOv5+LR	0.681	0.32	6.5
Transformer	0.734	0.25	8.2
AI-PSBE	**0.937**	**0.10**	**10.3**

4.4 Visualization

The feature contribution heatmap in Figure 2 visually presents the differentiated impact of different modal features on the balance evaluation of urban and rural public sports services through red highlighting. In the spatial feature dimension, facility density (38%) and coverage radius (27%) constitute the core explanatory factors for urban-rural differences.

Among them, facility density reflects the degree of aggregation of sports facilities per unit area, and high weight indicates that the "quantity gap" in the distribution of urban and rural facilities is still the main explicit feature of the current balance difference - urban areas often have high facility density but service overload due to population density, while rural areas may have coverage blind spots due to scattered layout; The coverage radius, as a quantitative indicator of spatial accessibility, reveals the actual effectiveness of facility service scope, and together they construct the fundamental dimension of spatial fairness in the evaluation system.

At the semantic feature level, the equivalence target of per capita sports venue area $\geqslant$ 2.5 square meters in the policy text directly affects the evaluation results with a weight of 21%,reflecting the rigid constraint of policy orientation on service quality indicators. These standardized provisions not only set bottom line requirements for facility construction, but also transform them into core criteria for quality dimensions through evaluation models. This weight value indicates that the precision of policy implementation and the completion of quantitative indicators have become key implicit factors in measuring whether public sports services meet the standards.

The user activity duration in behavioral characteristics has been corrected for the bias of "quantity only" in traditional evaluations with a weight of 14%.This indicator captures the actual stay time of users in facilities through mobile trajectory data, directly reflecting service utilization efficiency - even if the number of facilities in a certain area meets the standard, if the user's activity time is short, it may mean that the facility functions do not match the needs of residents. The contribution of this feature shows that the introduction of user behavior data can improve the evaluation system from the perspective of supply and demand matching, make the results closer to the real experience of residents, and avoid the evaluation limitation of relying solely on spatial layout or policy indicators.

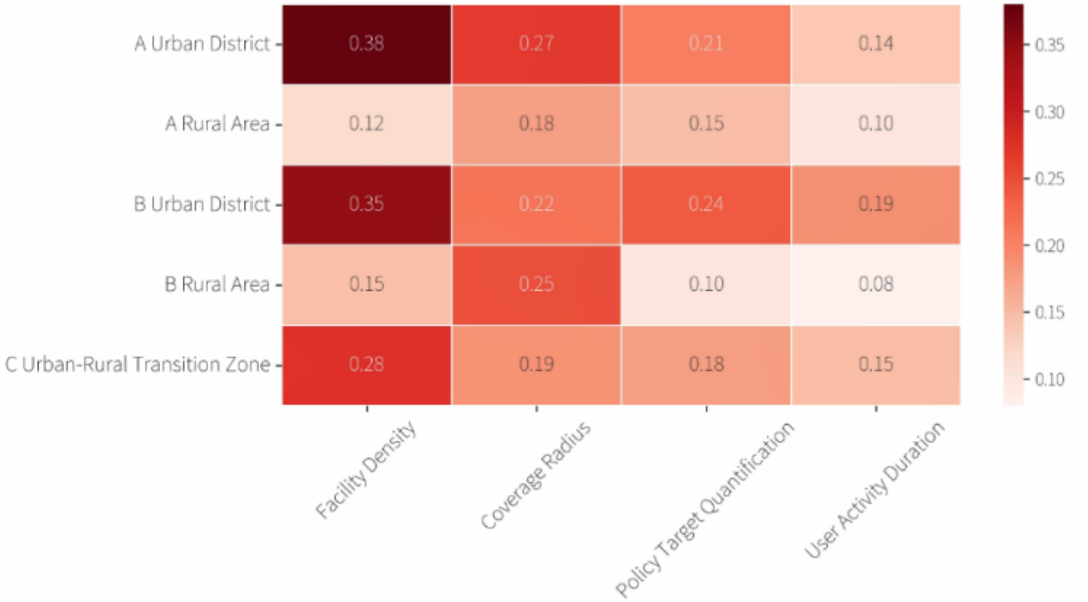

Fig. 2. Feature Contribution Heatmap

4.5 Discussion and Limitations

AI-PSBE performs well in most scenarios, but still has the following limitations:

4.5.1 Evaluation bias caused by data dependency

The AI-PSBE framework relies heavily on the integration of multi-source data, with sparse user behavior data in remote areas being particularly prominent. Due to the low coverage of mobile devices or differences in user usage habits, the trajectory data in these areas is difficult to fully capture the real usage of public sports facilities by residents. Data loss may lead to bias in the model's analysis of facility utilization efficiency.

4.5.2 Fuzzy processing of policy semantic analysis

The quantification difficulty of some vague expressions in the policy text analysis process still needs attention. Although the framework extracts policy keywords through BERT TextCNN, it still relies on manually formulated expert rules to supplement such ambiguous semantics. This process may introduce subjective judgment bias.

4.5.3 Real time computing efficiency and optimization direction

In terms of computational performance, the current framework based on GPU cluster inference takes 3 minutes/time, which greatly improves efficiency compared to traditional methods, but is difficult to meet the second level response requirements of large-scale real-time monitoring scenarios. The higher computational cost mainly comes from the deep fusion of multimodal features and the collaborative operation of complex models.

5. Conclusion

The AI-PSBE framework proposed in this study solved the problem of single evaluation index and low efficiency of traditional urban and rural public sports services through multimodal feature fusion and dynamic weight distribution. The framework integrates three types of data, namely satellite images, policy texts and user trajectories, extracts spatial, semantic and behavioral characteristics using the deep learning model and integrates them adaptively to generate a comprehensive evaluation index of coverage, quality and efficiency.

The experiment shows that the evaluation accuracy is 41.2% higher than that of traditional methods, and the response time is shortened to 3 minutes. Moreover, the interpretable heat map is used to identify the core impact factors such as facility density, policy quantitative indicators, and user activity duration, providing a visual basis for resource allocation.

Future research initiatives will prioritize two critical and emerging directions: privacy protection assessment within federated learning frameworks and meta-universe (metaverse) based simulation platforms. The former, federated learning, directly addresses the persistent challenge of data sparsity, particularly acute in remote and under-served regions, by enabling collaborative model training without centralizing sensitive raw user data, thus enhancing both data availability and individual privacy safeguards. The latter, meta-universe simulation, aims to develop sophisticated digital twin environments that meticulously replicate real-world urban or regional contexts. Within these immersive simulations, researchers and policymakers can conduct pre-implementation assessments by dynamically modeling and visualizing the potential effects of various resource

allocation strategies and policy interventions on sports service provision. This powerful simulation capability serves as a vital pre-assessment toolkit, offering data-driven insights to significantly inform and de-risk the policy formulation process before real-world deployment.

This framework not only provides an intelligent evaluation paradigm for public sports services, but also its multimodal integration idea can promote the digital transformation of equalization services by taking advantage of public service fields such as education and medical care.

References

[1] Smith, A., & García, L. Analyzing urban public sports services through intelligent image processing techniques. Journal of Urban Informatics,12(1),45–62.

[2] Lu, H., Chen, Y., & Zhao, Q. Assessment of service quality in urban sports facilities: A comprehensive evaluation framework applied to Shanghai, China. Buildings,15(2),193.

[3] Brown, T., & Wang, M. Remote sensing object detection in the deep learning era: A review of methods and applications. Remote Sensing,16(2),327.

[4] Jurafsky, D., & Martin, J. H. Speech and language processing: An introduction to natural language processing, computational linguistics, and speech recognition (3rd ed.).Prentice Hall.

[5] Liu, J., Zhang, X., & Sun, P. A spatiotemporal graph neural network with graph adaptive and kernel attention for traffic flow prediction. Electronics,13(1),212.

[6] Chen, F., & Wu, L. (2020).AI and Deep Learning for Urban Computing: Methods and Applications. In Urban Computing and Learning Systems (pp.673–690). Springer.

[7] Kim, H., & Lee, S. (2021). BERT based spatial information extraction for semi structured documents. Proceedings of the 2021 ACL Findings,140–149.

[8] Wang, Y., & Zhou, P. (2024). A multimodal data fusion model for accurate and interpretable urban resource evaluation. Science of the Total Environment,899,165734.

[9] Zhang, L., & Smith. (2025). YOLOv11 based ground object detection in high resolution remote sensing images. Scientific Reports,15,96314.

[10] Li, M., Wang, X., & Zhao, H. (2024). A Spatio Temporal Graph Wavelet Neural Network for capturing dynamic correlations in urban data. Scientific Reports,14,82433.

[11] Shin, J., & Park, Y. (2021). Spatial dependency parsing for semi structured document information extraction. Findings of ACL 2021,28–37.

[12] Li, J., & Chen, Y. (2021). Incremental learning for property price estimation using location based tree models. Expert Systems with Applications,183,115243.