Online-Journals.org

iJIM International Journal of Interactive Mobile Technologies

iJIM | eISSN: 1865-7923 | Vol. 19 No. 17 (2025) | OPEN ACCESS

https://doi.org/10.3991/ijim.v19i17.57859

PAPER

Innovative Applications of Augmented Reality and Interactive Mobile Technologies in Music Education

Shan Wang (✉)

School of General Education, Shanghai Technology and Innovation Vocational College, Shanghai, China

ws03130717@163.com

ABSTRACT

Against the backdrop of digital transformation in education, the integration of augmented reality (AR) and interactive mobile technologies offers new pathways for intelligent and personalized development in music education. While existing studies have explored the use of AR in musical learning environments, significant gaps remain in the dynamic capture of learner behavior data and its application in personalized adaptation. For instance, current AR-based music visualization systems often lack mechanisms for feedback on user behavior trajectories, and mobile learning platforms struggle to effectively integrate short- and long-term student behavior data in their recommendation modules. To address these limitations, this study proposes a music learning point-of-interest (POI) recommendation model based on spatiotemporal AR expansion and mobile interaction networks. The model operates through the synergistic functions of three components: a long-term behavior dependency module, a short-term behavior dependency module, and a balancing and integration module. Together, they enable precise content recommendation and context-aware learning path guidance. The findings of this study not only provide personalized learning support for music learners but also offer a technical framework and methodological reference for the development of personalized models in digital music education.

KEYWORDS

augmented reality (AR) technology, interactive mobile technology, music education, point-of-interest (POI) recommendation model, behavior dependency modules

1 INTRODUCTION

Amid the wave of digitalization sweeping across the education sector, music education is undergoing a profound transformation from traditional teaching models toward intelligent and personalized approaches [1–4]. Augmented reality (AR) technology, with its characteristic of seamlessly integrating virtual information with real-world scenes [5–7], and interactive mobile technology, with its capabilities of real-time connection and dynamic response [8, 9], provide new possibilities for breaking the limitations of time and space in music education and activating

Wang, S. (2025). Innovative Applications of Augmented Reality and Interactive Mobile Technologies in Music Education. *International Journal of Interactive Mobile Technologies (iJIM)*, 19(17), pp. 179–191. https://doi.org/10.3991/ijim.v19i17.57859

Article submitted 2025-03-22. Revision uploaded 2025-07-14. Final acceptance 2025-07-16.

learners' immersive experiences. This technological integration not only aligns with the requirements for teaching innovation in the era of Education Informatization 2.0 [10], but also responds to the core demands of practice-oriented, contextualized, and personalized cultivation in music education [11–13], becoming an important breakthrough point for promoting the upgrading of music education models.

Existing studies exploring the integration of AR technology and music education often ignore the dynamic capture of learner behavior data and its personalized adaptation [14–16]. For example, the music visualization system proposed in [17] only focuses on the 3D presentation of static music scores and fails to build a real-time feedback mechanism based on learners' behavior trajectories. The mobile music learning platform designed in [18], although introducing interactive functions, shows significant shortcomings in the integration and analysis of students' long-term behavior dependence and short-term behavior fluctuations in its point-of-interest (POI) recommendation module, leading to limited accuracy and adaptability of the recommendation results. These methodological limitations make it difficult for the technology applications to truly match the characteristics of significant individual differences among learners and the dynamic changes of learning contexts in the music learning process.

This paper focuses on constructing a music learning POI recommendation model based on AR spatiotemporal position expansion and mobile interaction networks. The model mainly consists of a student long-term behavior dependency module, a student short-term behavior dependency module, and a balancing and integration module. The long-term behavior dependency module mines learners' historical data in the music learning process to grasp their stable interest preferences and learning habits; the short-term behavior dependency module captures the current behavior dynamics and contextual changes of learners in real time; and the balancing and integration module achieves the organic integration and weight adjustment of the two types of behavior data through optimization algorithms, thereby realizing precise content recommendation and contextualized learning path guidance. The value of this study lies in the deep combination of AR's spatial positioning capability and the real-time data acquisition advantage of interactive mobile technology through technological integration and model innovation. It not only provides more targeted learning support for music learners but also offers a referable technical framework and methodology for the construction of personalized learning models in the field of digital music education, with important practical guiding significance for promoting the intelligent and precise development of music education.

2 MUSIC LEARNING POI RECOMMENDATION BASED ON AR SPATIOTEMPORAL POSITION EXPANSION AND MOBILE INTERACTION NETWORKS

2.1 Problem description and definitions

In order to clarify the problem of music learning POI recommendation, this paper first defines the overall check-in sequence, historical check-in sequence, and current check-in sequence.

Definition 1: Overall Check-in Sequence. In the music learning scenario based on AR spatiotemporal position expansion and mobile interaction networks, the overall check-in sequence is the global record of students' access trajectories to various POIs during the music learning process. For the student set in the AR mobile interaction network $I = \{i_1, i_2, \ldots, i_{|I|}\}$, the overall check-in sequence of each student i is denoted as $ZT = \{ZT_1, ZT_2, \ldots, ZT_{|v|}\}$, where v is the index of the current trajectory sequence.

Each ZT_u in the sequence consists of a set of music learning POIs $M = \{m_1, m_2, \ldots, m_{|M|}\}$ visited by student i in chronological order, such as $ZT_u = \{m_1, m_2, \ldots, m_{|ZT_u|}\}$. Each POI m is uniquely identified through AR-based spatiotemporal position expansion, combining the geographic coordinate tuple (LON_m, LAT_m) with AR scene anchor information, thus forming a complete set of music learning behavior trajectories that integrates physical space and AR virtual scenes.

Definition 2: Historical Check-in Sequence. The historical check-in sequence is a retrospective record of students' past music learning behaviors based on the overall check-in sequence. Given the overall check-in sequence of a student, the historical trajectory sequence of student i is defined as $\{ZT_1, ZT_2, \ldots, ZT_{|v-1|}\}$. This sequence covers all music learning POIs accessed by the student via the AR mobile interaction network before the current learning stage. It includes not only the geographic spatial location information of POIs but also the virtual teaching resource association data enabled by AR technology. The value of the historical check-in sequence lies in its ability to capture the student's long-term music learning preferences, knowledge acquisition trajectories, and habits in using AR technology, thereby providing data support for the "student long-term behavior dependency module."

Definition 3: Current Check-in Sequence. The current check-in sequence focuses on the recent behavioral dynamics of students in the AR mobile interaction network for music learning. For target student i, the current check-in sequence $ZT_v = \left\{m_1, m_2, \ldots, m_{|M_{s-1}|}\right\}$ represents the most recently accessed POIs within time stage s–1. Different from traditional check-in sequences, each POI mmm in this sequence is associated in real time with spatiotemporal contextual information generated by AR technology, such as AR-enabled virtual music interaction scenes triggered by the student at specific physical locations or real-time collaborative performance activities through mobile networks. The core function of the current check-in sequence is to reflect the student's immediate learning needs and contextual behavior fluctuations, providing real-time data input for the "student short-term behavior dependency module" to capture POI preference changes caused by course tasks, environmental shifts, or spontaneous interests in AR learning scenarios.

2.2 Recommendation model

The music learning POI recommendation model proposed in this paper is constructed based on students' check-in sequence data in the AR mobile interaction network and realizes accurate recommendation through the collaborative operation of three main modules: the student long-term behavior dependency module, the student short-term behavior dependency module, and the balancing and integration module. The student long-term behavior dependency module relies on the overall check-in sequence to deeply mine historical learning trajectories that integrate AR spatiotemporal anchor information and geographic encoding, capturing the stable interest preferences, AR usage habits, and knowledge mastery paths formed by students in long-term music learning. The student short-term behavior dependency module extracts dynamic features from the current check-in sequence and performs real-time analysis of behavioral fluctuations triggered by contextual factors such as AR virtual scene interaction and mobile collaborative learning, using operations such as time decay functions and AR contextual feature encoding. The balancing and integration module integrates the stable interests reflected in long-term behaviors and the immediate needs indicated by short-term behaviors through an adaptive weight adjustment mechanism. By combining the three-dimensional spatial position expansion information provided by AR technology and the real-time interaction data of mobile networks, a spatiotemporal context-aware POI recommendation algorithm is constructed to ultimately realize accurate prediction of the next music learning POI for students. Figure 1 presents the architecture of the music learning POI recommendation model.

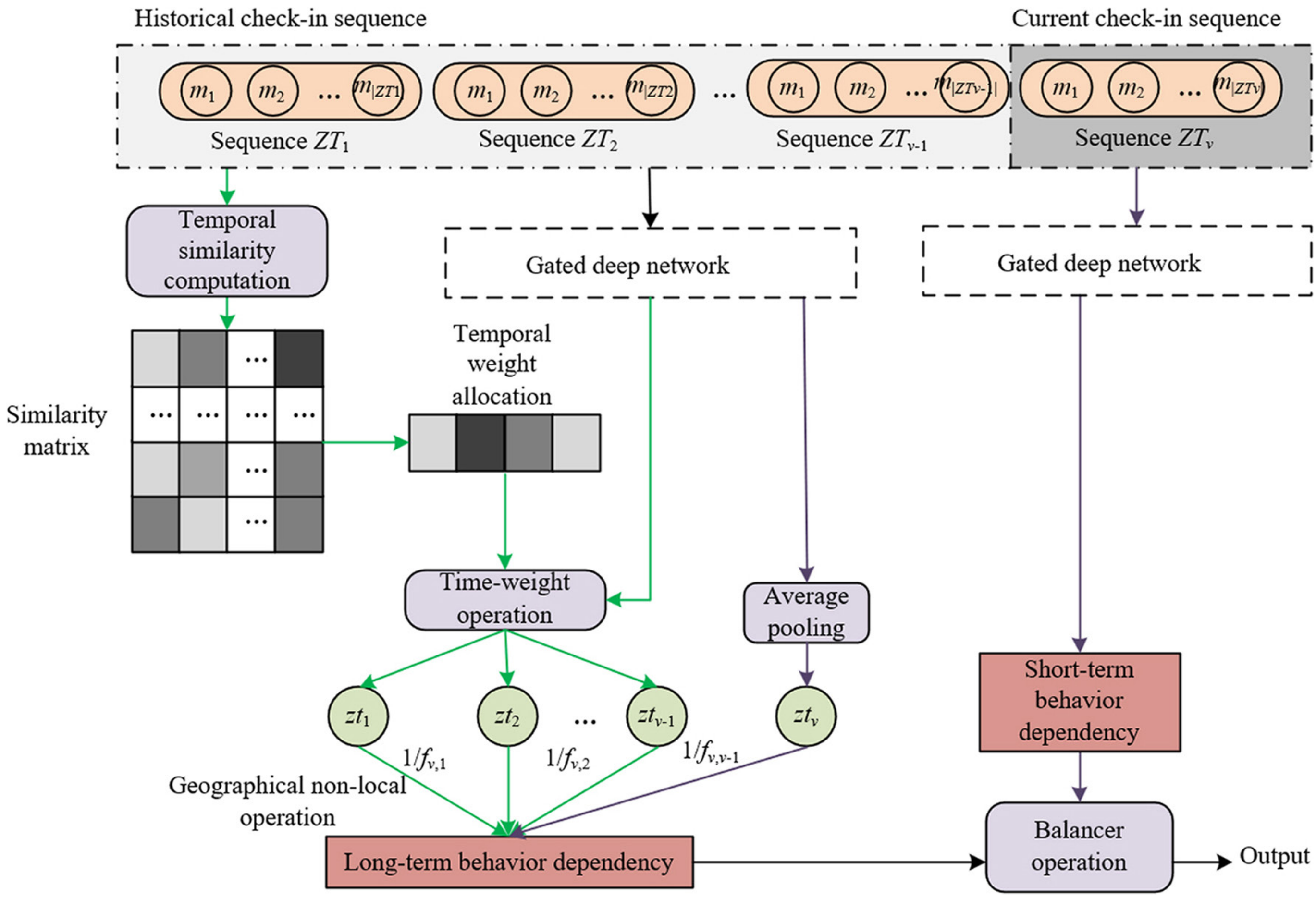

Fig. 1. Architecture of the music learning POI recommendation model

1. Student long-term behavior dependency module: Figure 2 shows the gated deep network architecture. This module is centered on a gated deep network, which consists of a gated RNN network and a collaborative network. The collaborative network adopts an LSTM architecture to perform temporal modeling on the overall check-in sequence recorded in the AR mobile interaction network. By capturing the sequence of music learning points-of-interest historically visited by students, it roughly extracts the stable interest patterns and knowledge evolution paths implicit in long-term learning behavior. Meanwhile, the collaborative network generates binary gate control signals based on Bernoulli distribution, which are used to regulate the attention distribution of the gated RNN network to the points-of-interest. When the binary gate output is 1, the gated RNN network performs deep encoding on the AR contextual features and geographic spatial information of the corresponding POI; when the output is 0, the influence of the POI is weakened, thus realizing the dynamic selection of key points-of-interest in long-term behavior and accurately mapping the overall interest preferences formed by students in AR music learning scenarios. Assume that the hidden states generated by the LSTM and RNN are represented as $g' = \left[g'_1, g'_2, \ldots, g'_s\right]$ and $g = [g_1, g_2, \ldots, g_s]$. The result through the fully connected layer and *sigmoid* layer is denoted as o_s. The binary gate formed by the Bernoulli distribution is denoted

as h_s, and the element-wise product is denoted by ⊗. The gated deep network is constructed by the following formulas:

$$\begin{aligned} g'_s &= LSTM\left(a_s, g'_{s-1}\right), s \in \left\{1,2,\ldots,\left|T_g\right|\right\} \\ o_s &= \text{sigmoid}\left(Ig'_s\right), h_s \sim \text{Bernoulli}(o_s) \end{aligned} \tag{1}$$

$$g_s = GRU(a_s, g_{s-1}), s \in \left\{1,2,\ldots,\left|T_g\right|\right\} \tag{2}$$

$$r_s = h_s \otimes g_s \tag{3}$$

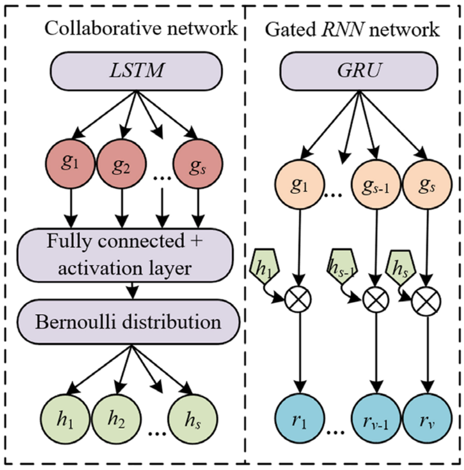

Fig. 2. Gated deep network architecture

In order to quantify the impact of spatiotemporal factors in AR environments on long-term behavior dependency, the module adopts a geo-global method to extract spatiotemporal features from the historical check-in sequence. First, the weekly check-in sequence is divided into 48 time slots according to weekdays and weekends. Combined with spatiotemporal factors, the historical sequence is encoded by the gated deep network as $\left\{r_1^{HIS}, r_2^{HIS}, \ldots, r_{\left|ZT_g\right|}^{HIS}\right\}$. For each time slot u, the set of points-of-interest visited by the student is constructed as $T_u = \left\{m_1, m_2, \ldots, m_{T_u}\right\}$, where each POI m_u contains (*LON*, *LAT*) geographic encoding and AR scene anchor information. By calculating the spatial similarity and temporal correlation of the points-of-interest within the time slot, a spatiotemporal feature matrix is constructed. This matrix is fused as a tensor with the POI feature vector output by the gated RNN network, so that the representation of long-term behavior includes not only the student's preference for music learning content, but also the spatiotemporal contextual information provided by AR technology, thereby capturing behavior dependency patterns under remote spatiotemporal dimensions. The similarity of each time slot can be calculated by the following formulas:

$$SIM_{u,k} = \frac{\left|T_u \cap T_k\right|}{\sqrt{\left|T_u\right| * \left|T_k\right|}}, (0 <= u < k <= 47) \tag{4}$$

$$zt_g = \sum_{s=1}^{\left|ZT_g\right|} Q_s r_s^{HIS}, Q_s = \frac{\exp\left(SIM_{z,T_s}\right)}{\sum_{u=1}^{\left|ZT_g\right|} \exp\left(SIM_{z,T_u}\right)} \tag{5}$$

The historical sequence is encoded as zt_g. Suppose that the historical sequence trajectory obtained by averaging the coordinate midpoints $\left(LON_{ZT_g}, LAT_{ZT_g}\right)$ is denoted by $ZT_g \in \{ZT_1, ZT_2, \ldots, ZT_{v-1}\}$. The geographic location encoding of the target student at the previous moment is denoted by $\left(LON_{m_{s-1}}, LAT_{m_{s-1}}\right)$. The geographic distance calculated using the haversine formula is represented by $f_{v,g}$. The geographic distance in the AR environment is denoted by $f_{v,g}$ as:

$$f_{v,g} = 2Rarc\_SIN\left(\sqrt{X+Y}\right), X = SIN^2\left(\frac{LON_{m_{s-1}} - LON_{ZT_g}}{2}\right),$$
$$Y = COS\left(LON_{ZT_g}\right)COS\left(LON_{m_{s-1}}\right)SIN^2\left(\frac{LAT_{m_{s-1}} - LAT_{ZT_g}}{2}\right) \tag{6}$$

After being processed by the gated deep network, the historical check-in sequence generates an encoded vector set $\left\{r_1^{CUR}, r_2^{CUR}, \ldots, r_{s-1}^{CUR}\right\}$, where each r_u^{CUR} corresponds to a fused feature of a historical POI. To avoid information loss, the module applies average pooling to aggregate the encoded vectors. This operation fuses the feature information of all historical points-of-interest with equal weight, which can not only preserve the stable interest trends in long-term behavior but also balance the feature fluctuations caused by AR spatiotemporal factors. For example, the behavior of a student accessing the same type of music knowledge point at different geographic locations through AR devices can, after average pooling, reinforce the interest weight of that knowledge point while preserving the differences in AR interaction details during each visit. The final generated long-term behavior representation vector will serve as the output of the "student long-term behavior dependency module," providing core features that reflect the student's overall interests for the balancing and integration module. The current check-in sequence is represented as zt_v as:

$$zt_v = \frac{1}{\left|ZT_v\right|}\sum_{s=1}^{\left|ZT_v\right|} r_s^{CUR} \tag{7}$$

Assume that the aggregation of the historical check-in sequence and the current check-in sequence is denoted as $\tilde{zt_v} = zt_g + zt_v$, and a trainable projection weight matrix is denoted by Q_g. The historical check-in sequence incorporating spatiotemporal factors is represented as zt_v^{LONG}.

$$zt_v^{LONG} = \frac{1}{\sum_g^{v-1} \exp\left(\frac{1}{f_{v,g}} \tilde{zt}_v^S zt_g\right)} \sum_g^{v-1} \exp\left(\frac{1}{f_{v,g}} \tilde{zt}_v^S zt_g\right) Q_g zt_g \tag{8}$$

2. Student short-term behavior dependency module: The student short-term behavior dependency module focuses on the dynamic feature extraction of the current check-in sequence. Aiming at the limitation of traditional RNNs in effectively handling the influence of spatial location, the module first quantifies the spatial correlation of points-of-interest in the AR mobile interaction network through a geographic matrix F. The geographic distance between the (LON, LAT) coordinates of each POI is calculated using the two-point distance formula to construct a distance filtering

mechanism: when the distance between POI m_u and m_k is less than a preset threshold, it is included in the non-continuous input sequence ZT_v^{GEO}, realizing jump sampling of spatially adjacent points-of-interest. The location expansion algorithm uses $j = 2$ as the initial jump length and dynamically adjusts the jump step to capture behavior fluctuations caused by spatial location changes in AR scenarios. For example, when a student accesses a virtual instrument library through an AR device in a music classroom, the algorithm prioritizes associating with other AR music learning resources in the same geographic area, embedding spatiotemporal contextual information into the short-term behavior representation.

Figure 3 shows the processing flow of the music learning POI sequence. At the sequence processing level, the idea of improved LSTM is referenced to construct a MogrifierGRU network, enhancing the interaction capability between GRU and contextual information to improve the accuracy of capturing short-term behavior dependency. A dual mapping mechanism is formed through linear transformation and ⊗ activation operations, which on the one hand performs temporal encoding of the points-of-interest in the current check-in sequence, and on the other hand fuses the 3D scene interaction data generated by AR technology with geographic coordinate features as a tensor. When e = 0, the network degenerates into the basic GRU mode, ensuring the extraction of traditional temporal features. By adjusting the jump length in the location expansion algorithm, the network can capture short-term behavior patterns at different spatial scales, such as student movement trajectories across different AR music learning sites on campus or remote collaborative learning behavior across regions. The finally generated short-term behavior representation zt_v^{SH} includes both the student's instantaneous learning interest fluctuations and the contextual features brought by AR spatiotemporal location expansion, providing dynamic decision support for the balancing and integration module.

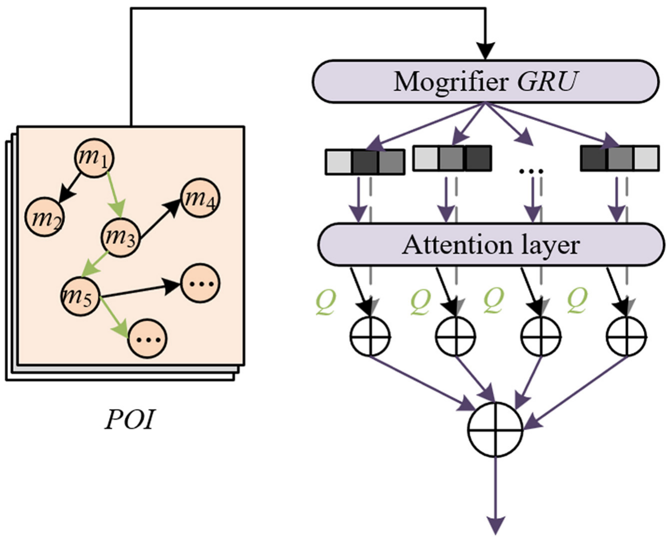

Fig. 3. Music learning POI sequence processing flow

3. Balancing and integration module: The balancing and integration module fuses AR spatiotemporal location expansion information and mobile interaction network behavior data as its basis. Through an adaptive weight mechanism, it dynamically balances the student's long-term behavior dependency representation zt_v^{Long} and short-term behavior dependency representation zt_v^{SH}. First, a spatiotemporal context-aware weight calculation model is constructed. Based on the AR scene anchor point information of the current check-in sequence and the

geographic heat distribution of the historical check-in sequence, a weight matrix reflecting spatiotemporal correlation is generated. Then, a gated fusion network is designed, which enhances long-term interest features related to the current AR learning scenario through an attention mechanism, while retaining the immediate needs generated by changes in AR context within the short-term behavior. Finally, the fused feature vector is input into a POI ranking generator. Combined with the 3D spatial location expansion information provided by AR technology and real-time collaborative data from the mobile network, a spatiotemporal-interest dual-dimensional recommendation ranking algorithm is constructed, realizing personalized ranking recommendation of music learning POI. This ensures that the recommendation results both align with the student's long-term interest development path and adapt to the immediate learning context in the AR environment. Assuming the probability set of each scene anchor point is denoted by o, and the trainable projection matrix is denoted by Q_e, the following expression is obtained:

$$o = \text{softmax}\left(Q_e\left(zt_v^{LONG} \oplus zt_v^{SH}\right)\right) \tag{9}$$

Assuming the total training samples are denoted by V, and the probability generated by the model for the ground truth POI of the j-th training sample is o_j, the objective function expression is:

$$LOSS = -\sum_{j=1}^{V} \log(o_j) \tag{10}$$

3 EXPERIMENTAL RESULTS AND ANALYSIS

From the data comparison of the recommendation performance of different models in Table 1, it can be seen that in both the training and testing sets, the proposed model demonstrates significant advantages in R@5, R@10, N@5, and N@10 indicators. In the training set, the R@5 of the proposed model reaches 0.3652 and R@10 is 0.4258; in the testing set, R@5 is 0.4125 and R@10 is 0.4789, all higher than the comparative models such as GRU, Bi-LSTM, and TCN. This indicates that the proposed music learning POI recommendation model effectively improves recommendation accuracy and relevance through the long-term behavior dependency module for mining stable interest preferences, the short-term behavior dependency module for capturing immediate dynamics, and the balancing and integration module for integrated regulation. It verifies the effectiveness of the method in the task of recommending music learning points-of-interest, providing learners with more accurate recommendations that better match their long-term habits and short-term contexts, supporting personalized learning in music education.

Table 1. Comparison of recommendation performance of different models

Model	Training Set				Test Set			
	R@5	R@10	N@5	N@10	R@5	R@10	N@5	N@10
GRU	0.0965	0.1147	0.0725	0.0812	0.2256	0.2563	0.1625	0.1856
Bi-LSTM	0.11256	0.1562	0.0879	0.1126	0.2136	0.2874	0.1586	0.1756
TCN	0.1458	0.1896	0.1126	0.1124	0.2789	0.3562	0.2236	0.2236
Attention-based RNN	0.1652	0.2236	0.1158	0.1326	0.2653	0.3321	0.1895	0.2156

(Continued)

Table 1. Comparison of recommendation performance of different models *(Continued)*

Model	Training Set				Test Set			
	R@5	R@10	N@5	N@10	R@5	R@10	N@5	N@10
ST-GCN	0.1236	0.1526	0.1126	0.1235	0.3215	0.3789	0.2156	0.2456
STPM	0.1589	0.2154	0.1236	0.1285	0.2653	0.3452	0.1852	0.2153
Time-LSTM	0.1896	0.2356	0.1358	0.1456	0.3154	0.4123	0.2356	0.2546
ST-ResNet	0.2456	0.3125	0.1879	0.1236	0.3356	0.4356	0.2456	0.2756
Proposed Model	0.3652	0.4258	0.3125	0.3258	0.4125	0.4789	0.3256	0.3268

Table 2. Ablation experiment results

Model	Training Set				Test Set			
	R@5	R@10	N@5	N@10	R@5	R@10	N@5	N@10
Position expansion algorithm with GRU	0.3569	0.4126	0.2896	0.3125	0.4123	0.4789	0.3265	0.3256
Gated deep network with LSTM	0.1896	0.2365	0.1325	0.1652	0.3256	0.4123	0.2351	0.2652
Without short-term behavior dependency	0.3452	0.3785	0.2635	0.2789	0.3895	0.4652	0.2789	0.3215
Without long-term behavior dependency	0.2236	0.2756	0.1789	0.2135	0.2456	0.3125	0.2135	0.2235
Excluding geographical non-local operation	0.3256	0.3652	0.2546	0.2756	0.3752	0.4452	0.2856	0.3289
Full model	0.3689	0.4256	0.3125	0.3269	0.4126	0.4756	0.3256	0.3251

From the ablation results in Table 2, the full model outperforms all variants in R@5, R@10, N@5, and N@10 metrics for both training and test sets. This indicates that the student long-term behavior dependency module mining stable interests, the student short-term behavior dependency module capturing immediate dynamics, and the inclusion of spatiotemporal-related operations such as geographic non-locality are effective. The cooperation of all modules allows comprehensive integration of long-term preferences and short-term contextual information, fully validating the scientific basis of the music learning POI recommendation method based on AR spatiotemporal position expansion and mobile interaction networks, and proving its precise capability in recommending suitable music learning interest points.

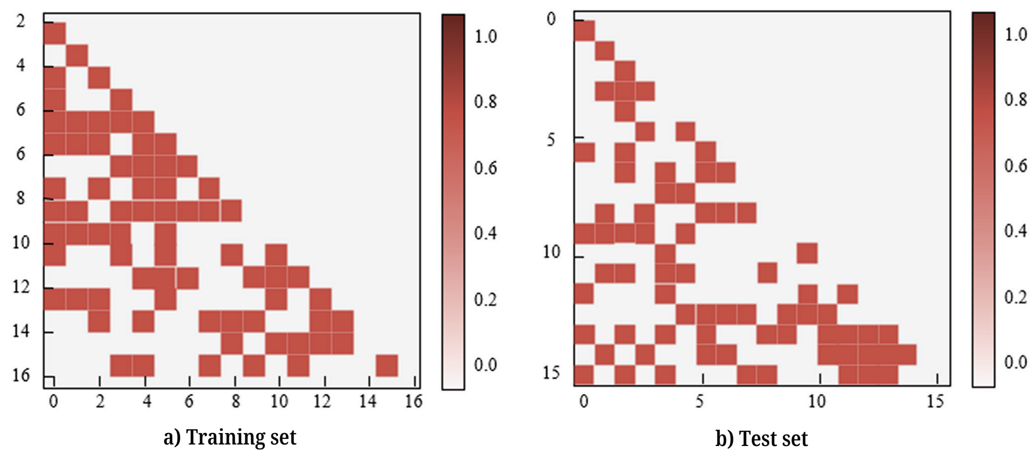

Fig. 4. Binary gate visualization

From the binary gate visualization results in Figure 4 for both training and test sets, the binary gate shows localized activation distribution patterns and does not exhibit globally uniform or fully covering weight associations. This indicates that our model does not apply global weighting such as general attention mechanisms. Moreover, these activation patterns correspond to different interest-point associations in the student check-in sequences. During the extraction by the student long-term behavior dependency module of historical data and the capture by the short-term behavior dependency module of current dynamics, the collaborative network, relying on this binary gating mechanism, captures local dependency relationships among interest points in the check-in sequence. From training to test sets, although the activation patterns vary, they consistently reflect behavioral dependency characteristics. This verifies that the collaborative network helps discover behavior dependencies in user check-in sequences. The model's unique design of non-global weight consideration combined with the collaborative network is effective in adapting to AR music education scenarios and capturing behavioral dependencies, laying the foundation for precise recommendation of learning interest points.

From the performance comparisons in Figure 5, our AR position expansion algorithm outperforms other baseline models in both training and test sets on metrics such as NDCG@10, NDCG@5, Rec@10, and Rec@5. In the training set, our method achieves higher Rec@10 than models such as TCN, GASC-Net, and HMM. In the test set, it also excels in NDCG@5 and Rec@5. Aligned with the research content, this advantage stems from the long-term behavior dependency module's mining of stable preferences from historical data and the short-term behavior dependency module's capture of current behavior dynamics and contextual changes. Through the balancing and integration module, the two types of behavior data are organically combined and weighted. This synergy enables our method—based on AR spatio-temporal position expansion and mobile interaction networks—to more accurately understand student music learning needs and surpass traditional baselines in recommendation performance, strongly validating its effectiveness in personalized recommendation scenarios in music education.

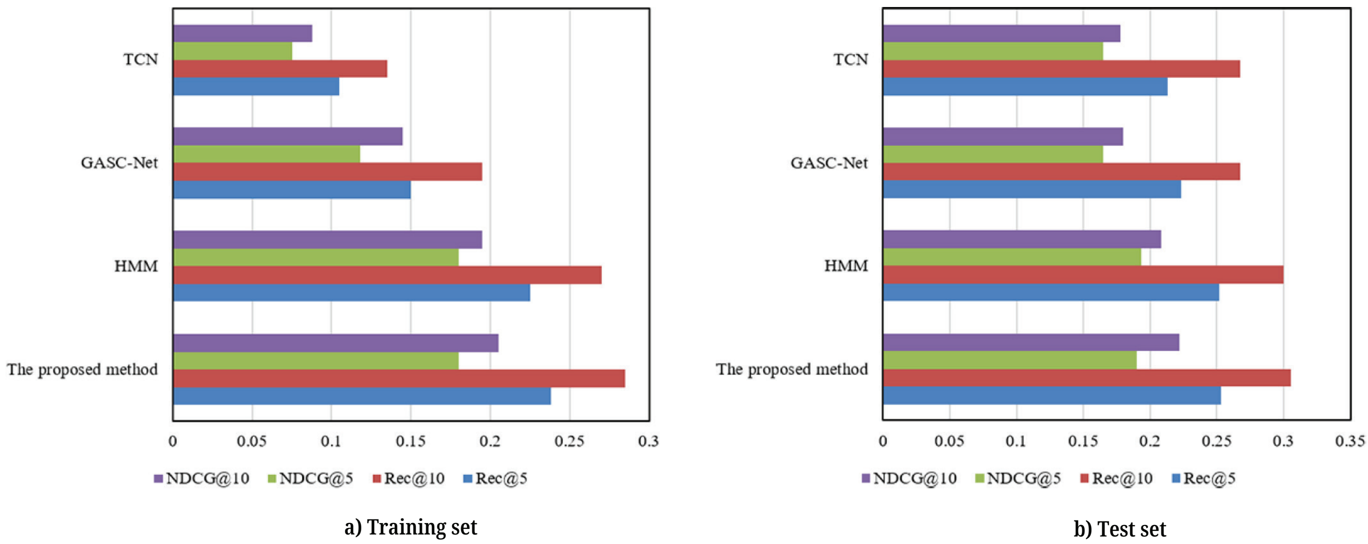

Fig. 5. Performance comparison of AR position expansion algorithm and other baseline models

4 CONCLUSION

This study, focusing on "innovative applications of AR and interactive mobile technologies in music education," constructs a music learning POI recommendation model based on AR spatiotemporal position expansion and mobile interaction networks, with significant value. On the technical application level, it innovatively integrates AR and mobile interaction networks, breaks through traditional recommendation models in music education, deeply embeds spatiotemporal location information into the mining of learning interest points, and provides a new path for digital transformation in music education. In terms of model effectiveness, through multiple comparative experiments, the model—supported by long-term behavior dependency for stable interest mining, short-term behavior dependency for capturing immediate dynamics, and the synergy of the balancing and integration module—performs excellently on metrics such as R@5, R@10, and NDCG. It can accurately adapt to students' long-term learning habits and short-term contextual needs, effectively improving recommendation accuracy and personalization of music learning interest points and constructing more fitting contextualized learning paths for learners, contributing to the learner-centric teaching transformation in music education.

Although the study achieves results, limitations remain. On one hand, the model heavily depends on AR environments; in scenes with limited AR device performance, data collection and processing efficiency decline, possibly affecting recommendation timeliness. On the other hand, the behavioral data mining dimension focuses relatively on spatiotemporal and interaction sequences, lacking consideration of complex factors such as student musical learning emotion and social collaboration, making it difficult to fully represent deep learner needs and limiting model adaptability to complex learning contexts. Future research can advance in three aspects. On technical adaptation, optimize model compatibility and lightweight processing for heterogeneous AR devices, integrate edge computing to reduce device performance impact on recommendation services and ensure real-time recommendation experience. On data dimension expansion, introduce affective computing and social network analysis technologies to collect and analyze students' emotional states and social interaction relationships in AR music learning, enriching behavioral data features and enhancing model capture of diverse needs. On application scenario extension, explore model adaptability optimization in different music education stages and different AR interaction modes, promoting deep implementation and continuous innovation of AR technology in music education recommendation scenarios, and supporting the construction of a more intelligent and immersive music education ecosystem.

5 ACKNOWLEDGEMENT

This study is supported by the Shanghai Educational Science Research Project "Theoretical and Practical Research on Constructing a Multidimensional Evaluation System for Ideological and Political Education in Aesthetic Education Courses under the 'Double High Plan' Initiative" (Grant No. C2025313), and the Key Project of Shanghai Vocational Education Association "Research on the Development and Implementation Pathways of Aesthetic Education Courses in Higher Vocational Colleges from the Perspective of Cultivating Virtue and Nurturing Talents" (Grant No. ZD202401).

6 REFERENCES

[1] Y. Zhang, B. W. Fen, C. Zhang, and S. Pi, "Transforming music education through artificial intelligence: A systematic literature review on enhancing music teaching and learning," *International Journal of Interactive Mobile Technologies (iJIM)*, vol. 18, no. 18, pp. 76–93, 2024. https://doi.org/10.3991/ijim.v18i18.50545

[2] N. Sun and Y. Zang, "Innovative applications and teaching effectiveness analysis of interactive mobile technology in music education," *International Journal of Interactive Mobile Technologies (iJIM)*, vol. 19, no. 1, pp. 93–106, 2025. https://doi.org/10.3991/ijim.v19i01.53497

[3] S. Özçelik and N. Hardalaç, "The statistical measurements and neural network analysis of the effect of musical education to musical hearing and sensing," *Expert Systems with Applications*, vol. 38, no. 8, pp. 9517–9521, 2011. https://doi.org/10.1016/j.eswa.2011.01.149

[4] O. Shevtsova, V. Tsarenko, S. Kurkina, P. Voloshyn, and T. Lisovska, "The importance of musical and aesthetic education of young people in modern society," *Amazonia Investiga*, vol. 12, no. 61, pp. 51–60, 2023. https://doi.org/10.34069/AI/2023.61.01.6

[5] J. Grubert, T. Langlotz, S. Zollmann, and H. Regenbrecht, "Towards pervasive augmented reality: Context-awareness in augmented reality," *IEEE Transactions on Visualization and Computer Graphics*, vol. 23, no. 6, pp. 1706–1724, 2016. https://doi.org/10.1109/TVCG.2016.2543720

[6] Z. Nurbekova and B. Baigusheva, "On the issue of compliance with didactic principles in learning using augmented reality," *International Journal of Emerging Technologies in Learning (iJET)*, vol. 15, no. 15, pp. 121–132, 2020. https://doi.org/10.3991/ijet.v15i15.14399

[7] B. Baran, E. Yecan, B. Kaptan, and O. Paşayiğit, "Using augmented reality to teach fifth grade students about electrical circuits," *Education and Information Technologies*, vol. 25, pp. 1371–1385, 2020. https://doi.org/10.1007/s10639-019-10001-9

[8] M. Gerla, D. Maggiorini, C. E. Palazzi, and A. Bujari, "A survey on interactive games over mobile networks," *Wireless Communications and Mobile Computing*, vol. 13, no. 3, pp. 212–229, 2013. https://doi.org/10.1002/wcm.2197

[9] C. Dibak, B. Haasdonk, A. Schmidt, F. Dürr, and K. Rothermel, "Enabling interactive mobile simulations through distributed reduced models," *Pervasive and Mobile Computing*, vol. 45, pp. 19–34, 2018. https://doi.org/10.1016/j.pmcj.2018.02.002

[10] A. Grafstein, "A discipline-based approach to information literacy," *The Journal of Academic Librarianship*, vol. 28, no. 4, pp. 197–204, 2002. https://doi.org/10.1016/S0099-1333(02)00283-5

[11] R. Cremades and L. Herrera, "A comparative study of formal and informal education in musical knowledge of students of professional education in music," *Revista Publicaciones*, vol. 40, pp. 73–87, 2010.

[12] C. R. Bellochio, L. W. F. Garbosa, and L. F. Lazzarin, "The education and arts research line and the studies in musical education in the context of the Post-Graduation Program in Education – UFSM," *Educacao*, vol. 37, no. 1, pp. 13–30, 2012. https://doi.org/10.5902/198464444111

[13] P. Marcheva, "Variations of the learning content in music about expanding students' musical experience in the intercultural education," *Pedagogika-Pedagogy*, vol. 86, no. 1, pp. 30–40, 2014.

[14] F. Dirgenali and S. Kara, "Comparison of an eigenvector music method with classical FFT and model-based AR methods in the analysis of mitral valve Doppler signals," *Instrumentation Science and Technology*, vol. 33, no. 5, pp. 575–595, 2005. https://doi.org/10.1080/10739140500226684

[15] O. E. Bolívar-Chávez, J. Paredes-Labra, Y. V. Palma-García, and Y. A. Mendieta-Torres, "Educational technologies and their application to music education: An action-research study in an Ecuadorian university," *Mathematics*, vol. 9, no. 4, p. 412, 2021. https://doi.org/10.3390/math9040412

[16] B. Mei and S. Yang, "Chinese pre-service music teachers' perceptions of augmented reality-assisted musical instrument learning," *Frontiers in Psychology*, vol. 12, p. 609028, 2021. https://doi.org/10.3389/fpsyg.2021.609028

[17] O. Lopez-Rincon and O. Starostenko, "Music visualization based on spherical projection with adjustable metrics," *IEEE Access*, vol. 7, pp. 140344–140354, 2019. https://doi.org/10.1109/ACCESS.2019.2944083

[18] M. Paule-Ruiz, V. Álvarez-García, J. R. Pérez-Pérez, M. Álvarez-Sierra, and F. Trespalacios-Menéndez, "Music learning in preschool with mobile devices," *Behaviour & Information Technology*, vol. 36, no. 1, pp. 95–111, 2017. https://doi.org/10.1080/0144929X.2016.1198421

7 AUTHOR

Shan Wang received the Master degree from Nanjing University of the Arts, P.R. China. Now, she is a Lecturer at Shanghai Technology and Innovation Vocational College. Her research focuses on musicology and theater and traditional Chinese opera studies (E-mail: ws03130717@163.com).