Online-Journals.org

iJIM International Journal of
Interactive Mobile Technologies

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

https://doi.org/10.3991/ijim.v19i19.58321

PAPER

Design of an Adaptive Learning System for Music Education Empowered by Interactive Mobile Technologies

Kai Zhang (✉)

Cai Yuanpei School of Art and Design, Shaoxing University, Shaoxing, China

18657598885@163.com

ABSTRACT

With the widespread adoption of mobile devices and the expansion of 5G networks, interactive mobile technologies have broken the temporal and spatial constraints of traditional music education. Learners can now access diverse educational resources and engage in flexible interactions via mobile platforms. However, current recommendation systems for music education primarily rely on basic label matching, which fails to accommodate learners' varying performance levels, learning paces, and the dynamic demands of mobile contexts. This limitation hinders the progress of digital transformation in music education. Existing approaches suffer from significant drawbacks: collaborative filtering models focus solely on binary learner-resource interactions and overlook the impact of multi-directional interactions; content-based recommendations emphasize musical attributes of resources but ignore real-time mobile interaction features; and graph neural network (GNN) models lack optimization for progressive music skill development and real-time feedback in practice scenarios, resulting in suboptimal alignment between recommendations and learning trajectories. To address these challenges, this study proposes a music education resource recommendation model based on a multi-directional mobile interaction graph attention network (GAN). The main innovations of this research include: (1) integrating multi-directional mobile interaction features with spatiotemporal dynamics to overcome the limitations of single-interaction modeling in traditional systems; (2) combining GAN with mobile interaction-aware modules to dynamically capture interaction weights and adapt to real-time learning contexts; and (3) embedding principles of progressive musical skill development into the learning algorithm to enhance alignment between recommended resources and learners' advancement. This study provides a technical foundation for adaptive learning systems empowered by interactive mobile technologies, promoting greater precision and personalization in music education resource recommendation.

KEYWORDS

interactive mobile technology, music education, adaptive learning system, resource recommendation model, graph attention network (GAN), multi-directional interaction features

Zhang, K. (2025). Design of an Adaptive Learning System for Music Education Empowered by Interactive Mobile Technologies. *International Journal of Interactive Mobile Technologies (iJIM)*, 19(19), pp. 55–69. https://doi.org/10.3991/ijim.v19i19.58321

Article submitted 2025-04-16. Revision uploaded 2025-07-20. Final acceptance 2025-07-22.

1 INTRODUCTION

With the popularization of mobile terminals such as smartphones and tablets [1–3] and the widespread coverage of 5G networks [4, 5], interactive mobile technology is profoundly reshaping the form of music education [6–9]. The portability of mobile terminals breaks the temporal and spatial constraints of traditional classrooms [10–13]. Learners can access diversified resources such as sheet music, audio, and video through mobile devices. Technologies such as touch interaction and voice interaction also make music practice and teacher-student interaction more flexible [14, 15], for example, learners can annotate difficult points on sheet music and give voice feedback about their confusion, while teachers can correct assignments remotely. However, current music resource recommendations mostly rely on simple label matching [16], which is difficult to adapt to learners' performance level, learning pace, and the dynamic demands of mobile scenarios. This coarse-grained model easily dampens learning enthusiasm and wastes fragmented time. Therefore, using interactive mobile technology to construct adaptive learning systems that precisely respond to individual differences has become an important topic in the digital transformation of music education.

Existing studies show obvious limitations in learning resource recommendation. The collaborative filtering model in literature [17] only focuses on the binary interaction between student and resource, ignoring the influence of multi-directional interactions such as teacher-student and peer-peer on learning needs. For example, it cannot adjust recommendations based on the teacher's evaluation of "need to strengthen rhythm training." The content-based recommendation in literature [18] emphasizes the musical attributes of resources but does not include real-time mobile interaction features such as practice duration and device location, which may result in recommending long video resources during commuting scenarios. Although the GNN model in literature [19] considers relationship modeling, it is not optimized for scenarios such as progressive music skill development and real-time feedback in practice. It may skip basic resource recommendations and push advanced content or fail to adjust recommendations based on pitch deviation and other real-time errors, leading to insufficient matching with the learning process.

The core of this paper is to construct a music education resource recommendation model based on a multi-directional mobile interaction GAN, and it develops from three aspects: at the feature representation level, it extracts multi-directional interaction features between learners and resources, teachers, and peers, and integrates spatiotemporal dynamic features of mobile scenarios; at the model architecture level, it uses a GAN to capture the weight differences of interaction relationships and designs a mobile interaction perception module to adjust recommendation strategies in real time; at the learning algorithm level, it designs a loss function based on the progression rules of musical skills. Its value lies in: technically, it compensates for the deficiency of existing models in capturing multi-directional mobile interaction features, improving recommendation accuracy and scenario adaptability; in application, it enables learners to efficiently obtain suitable resources and improve learning outcomes; and at the industry level, it provides a feasible solution for adaptive systems and promotes the digital transformation of music education toward personalization and precision.

2 MUSIC EDUCATION RESOURCE RECOMMENDATION BASED ON MULTI-DIRECTIONAL MOBILE INTERACTION GAN

The core objective of this paper is to rely on the real-time, portable, and interactive advantages of interactive mobile technology to construct an adaptive learning system that can precisely match individual differences among learners, ultimately realizing the goals of personalization, efficiency, and scenario-based empowerment in music education. Advanced music education resource recommendation methods serve as the key link between interactive mobile technology and this core objective. On one hand, music education resources have strong professionalism and high heterogeneity, and with the support of mobile technology, the number of resources has grown explosively. Without scientific recommendation methods, learners can easily fall into the dilemma of "resource overload," making it difficult to efficiently obtain suitable content. On the other hand, the core of adaptive learning lies in "dynamic adaptation," requiring the system to continuously adjust resource provision based on learners' real-time interaction data. Advanced recommendation methods can accurately model learner characteristics and dynamically optimize resource matching logic. They can not only use the real-time data collection capability of mobile technology to capture learning state changes but also push adaptive resources precisely to mobile terminals, directly supporting the implementation of adaptive learning scenarios of "on-demand recommendation and dynamic adjustment."

The following sections will provide a detailed explanation of the music education resource recommendation model based on the multi-directional mobile interaction GAN from aspects such as feature representation and model architecture.

2.1 Feature representation

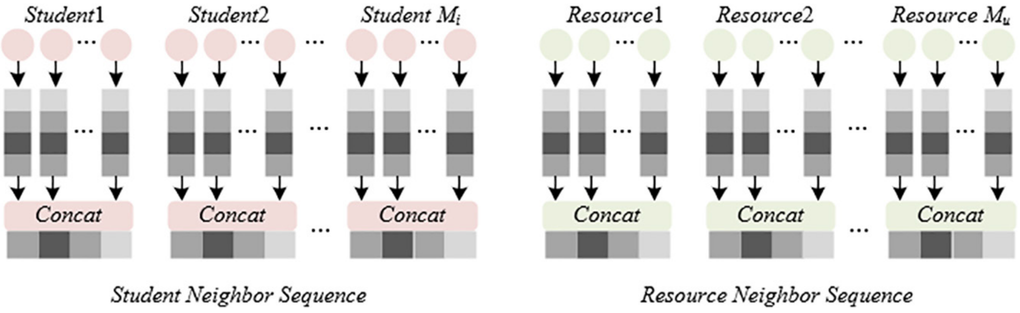


Fig. 1. Schematic diagram of student and music resource neighbor sequences

The basic principle of feature representation in the model is based on the structured modeling of dynamic interaction relationships. The core is to achieve precise extraction and ordered expression of multi-dimensional features through the student–music education resource mobile interaction graph $h = (N_i, N_u, \gamma)$. Among them, N_i and N_u represent the student node set and the music education resource node set, respectively; γ is the edge set. For any $(i, u) \in \gamma$, there are $n_i \in N_i$ and $n_u \in N_u$. The edge set γ consists of all edges of the form (n_i, n_u, p), where n_i is the student node, n_u is the music education resource node, and interaction has occurred between student i and resource u. Label p indicates the interaction order. The core goal of this model is to "capture the dynamic interaction pattern and adapt to individual student differences and mobile scenario needs," and four groups of categorical features are taken as the basic sources of feature representation: student attributes and music education resource attributes constitute the basic features, directly reflecting the inherent attributes of students and resources, while the student neighbor sequence and music education resource neighbor sequence are generated based on the concept of ordered neighbors in the dynamic interaction graph. The student neighbor sequence is composed of the resource attribute features of the student's ordered neighbors. For example, if a student first interacts with "beginner piano etudes" and then "intermediate piano etudes," the neighbor sequence integrates the difficulty, rhythm, and other attributes of these two types of resources in this order. The music education resource neighbor sequence is composed of the student attribute features of the resource's ordered neighbors. For example, if a "violin beginner tutorial" is first interacted with by "beginner *A*" and later by "intermediate learner *B*," then its neighbor sequence integrates the learning level, practice frequency, and other attributes of the two students in that order. Figure 1 shows the schematic diagram of student and music recommendation resource neighbor sequences. Assume that the total number of interactions of n is represented by M, and each node in V_n is called the neighbor of n, then:

$$\begin{aligned} & V_n = (n_1, n_2, \ldots, n_M) \\ & s.t.\ \ \forall 1 \le m \le M, (n, n_m, p_m) \in \gamma \ \ OR \ \ (n_m, n, p_m) \in \gamma \\ & \forall 1 \le k < j \le M, p_k < p_j \end{aligned} \tag{1}$$

The interaction graph h changes over time. When encountering new students and new music education resources, new nodes are continuously inserted into V_i or V_u and newly occurring interactions are inserted into γ. The interaction graph at time s is denoted as h^s. This feature extraction method assigns temporal attributes to neighbor sequences through edge labels so that features not only contain static attributes but also reflect the dynamic evolution trajectory of the "student–resource" interaction, providing structured feature support for capturing student learning progress and changes in the resource adaptation group.

2.2 Model architecture

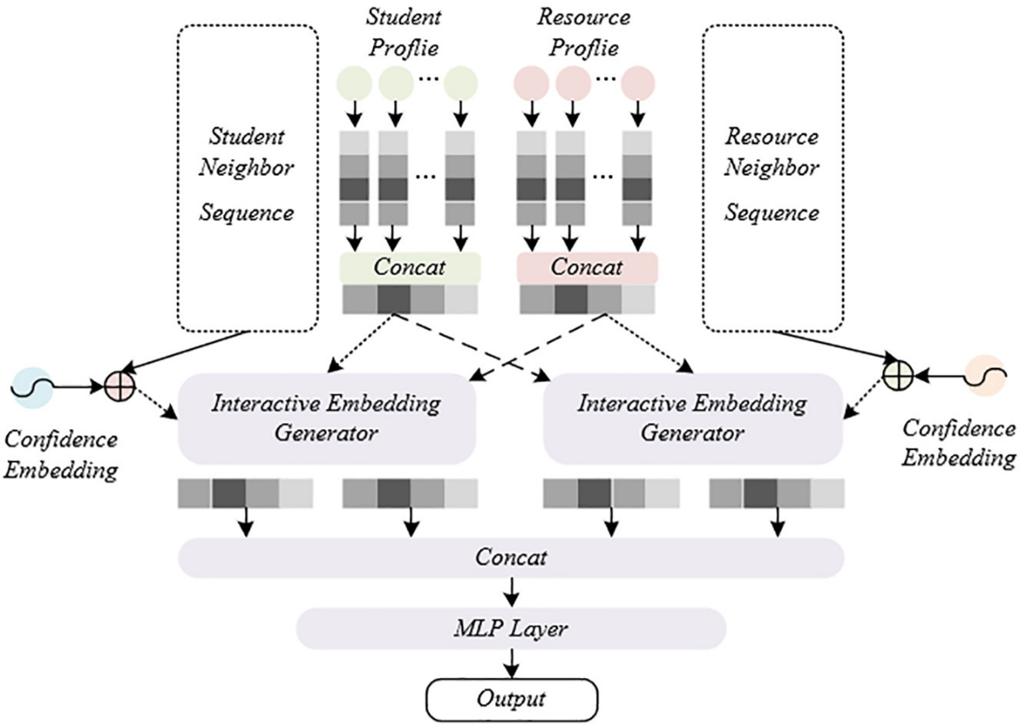


Fig. 2. Model architecture

The basic architecture of the model realizes feature processing and recommendation prediction through the collaboration of four parts. The specific architecture is shown in Figure 2. The first part is the embedding layer, which is responsible for converting sparse features such as student attributes and music education resource attributes into low-dimensional dense embedding vectors and mapping discrete interaction information in student neighbor sequences and music education resource neighbor sequences into computable vector representations, laying the foundation for subsequent feature fusion. The second part is the confidence embedding vector. For edges marked in order in the student–music education resource mobile interaction graph, differentiated embedding vectors are assigned to interactive neighbors at different positions. For example, the confidence embedding weight of recent interactions is higher than that of earlier interactions, and the confidence embedding weight in stable scenarios is higher than that in noisy scenarios, so as to quantify the reliability of different interactions and avoid interference from accidental or outdated interactions in recommendations. The third part is the mobile interaction embedding generator, which serves as the core module of the architecture and performs feature aggregation on the student–resource interaction graph based on the GAN: first, the vectors output from the embedding layer are used as the initial node features, and the attention mechanism is used to calculate the interaction weights between students and resources; then, the aggregation weights are adjusted with the confidence embedding vectors, and the node association strength is dynamically updated according to the interaction order; finally, two types of key features are generated—graph node feature representations based on mobile interaction information and adaptive historical mobile interaction feature representations—thus realizing deep modeling of the “student–resource” multi-directional interaction relationship. The fourth part is the multi-layer perceptron (MLP) layer, which fuses the two types of features output by the mobile interaction embedding generator with the basic attribute vectors from the embedding layer and learns the complex associations among features through multiple layers of nonlinear transformation, finally outputting the student’s preference probability for a specific music education resource.

The core construction goal of the embedding layer is to convert the sparse features of students and music education resources into computable dense vectors, laying the foundation for the GAN to capture interaction relationships, and this design directly serves the research objective of “accurately capturing mobile interaction features to achieve personalized resource recommendation.” Specifically, the model realizes feature transformation by constructing two basic embedding tables: the student embedding table $R_i = [r_{i1}, r_{i2}, \ldots, r_{iVi}]$ consists of several embedding vectors with dimension G_i, corresponding to all categorical attributes of the student, mapping the originally discrete and sparse student attributes into low-dimensional dense vectors; the music education resource embedding table R_i also consists of embedding vectors with dimension G_i, corresponding to the resource’s categorical attributes, similarly transforming the sparse features of the resource into dense vectors. This “sparse-to-dense” design solves the problem of high dimensionality and sparse distribution in original features, enabling the model to perceive the attribute features of students and resources through vector operations while retaining the core information of the features. The deep construction logic of the embedding layer is reflected in the design of shared embedding tables, which enhances the interaction semantics between students and resources through feature association, aligning with the research requirement of “modeling dynamic relationships based on mobile interaction graph.” The student neighbor sequence shares the embedding table R_u with the music education resource attributes, meaning that the vector representations of resources previously interacted with by the student and the attribute vectors of these resources come from the same embedding space. This allows the model to directly associate the student's interaction history with the inherent attributes of the resources. Similarly, the music education resource neighbor sequence shares the embedding table R_i with the student attributes, allowing the group of students who interacted with the resource to form vector associations with the resource's attributes. This shared mechanism breaks the separation between attribute features and interaction features, making the embedding vectors simultaneously carry the dual semantics of "individual attributes" and "interaction relationships," providing a richer feature basis for the GAN to capture interaction weights.

The core design goal of the confidence embedding vector is to accurately capture the temporal dimension features of mobile interaction, and by quantifying the confidence differences of interactions at different time points, it provides a "time-aware" feature basis for the GAN, ultimately serving the recommendation goal of "adapting to students' dynamic learning preferences." Its design directly responds to the principle in music education that "recent interactions better reflect current needs." For example, the piano pieces practiced by a student this week more accurately reflect their current performance level than those practiced last month, so the model needs to strengthen the weight of recent interactions through confidence differentiation. Suppose the interaction order is represented by m, the current unit's position index in the embedding vector is represented by u, the total number of interactions in the sequence is represented by M, and the dimension of the embedding vector is represented by G, then the initialization of the confidence embedding vector is formalized as:

$$ZR_{(m,u)} = \exp(m - M - 1)COS((u - 1)\tau/G) \tag{2}$$

According to the above formula, the exponential part of the initialization formula models the trend that "confidence decays as the time interval increases," allowing recently occurred mobile interactions to obtain higher confidence weights, while earlier interactions naturally have lower weights. This mechanism precisely aligns with the dynamic changes in student preferences during music learning. The cosine part, based on the theory of discrete Fourier transform, uses a half-period cosine curve to depict the sequential order of interactions. For example, the sequence in which a student first practices "*C* major basic scales" and then moves on to "*C* major simple pieces" is transformed into computable vector features through the periodic changes of the cosine curve. This allows the model to perceive the progression of the learning path and avoids treating disordered interactions equally. This dual design of "decay + sequence" enables the confidence embedding vector to distinguish between the recency of interactions and capture the logical order of interactions, providing a time-dimensional judgment basis for the subsequent attention mechanism to calculate importance.

The deep design logic of the confidence embedding vector lies in the "fusion of temporal features and interaction features," which enhances the ability to capture dynamic relationships in the mobile interaction graph by adapting to the modeling needs of the GAN. Its design draws on the idea of positional encoding in neural machine translation but is further optimized for the music education scenario: positional encoding only distinguishes sequence positions, while the confidence embedding vector additionally models "confidence decay" through an exponential factor. In music education, this means not only knowing "student learns *A* before *B*" but also clarifying that "the interaction with *B* is more important for the current recommendation than the interaction with *A*." In the model process, the confidence embedding vector is added dimension-wise to the mobile interaction neighbor embedding sequence processed by the embedding layer so that the neighbor embeddings simultaneously carry three types of information: resource attributes, interaction relationships, and temporal confidence. This fusion allows the GAN to naturally incorporate the time factor when calculating interaction weights between students and resources. For example, when aggregating features of the student neighbor sequence, the feature of a recently interacted resource will be assigned higher attention weight due to higher confidence, thereby making the recommendation results closer to the student's current learning state and ultimately improving the model's response accuracy to dynamic learning needs in mobile scenarios.

The core construction goal of the mobile interaction embedding generator is to finely process interaction features to accurately portray students' dynamic interests and the attractiveness of music education resources. This design directly serves the research objective of "improving the adaptability of the recommendation ranking model to dynamic interactions." Its multi-directional attention layer, as a core component, takes the student attribute, resource attribute, student neighbor sequence, and resource neighbor sequence embedding vectors output from the embedding layer as input, and realizes deep analysis of interaction relationships through a dual attention mechanism: on one hand, it calculates mobile interaction weights to measure the importance of each neighbor to the current node. For example, a resource like "piano scale practice" that a student frequently practiced recently is significantly more representative of current interest than a "music theory knowledge" resource browsed casually earlier. On the other hand, it calculates adaptive weights to evaluate the relevance of neighbors to the recommendation context. For example, when a student is in the "rhythm training" phase, resource neighbors labeled with "rhythm practice" have higher relevance weights in recommendation. This dual-weight design breaks through the traditional model's simple aggregation of interaction features, allowing attention allocation to match historical interaction patterns and respond to real-time recommendation scenarios, thus providing semantically directed interaction features for subsequent feature aggregation. Specifically, r_i, r_u, r_{ivt}, and r_{uvt} respectively denote the student attribute, music education resource attribute, student neighbor sequence, and music education resource neighbor sequence. These embedding vectors are used as input to the multi-directional attention layer. Assume four independent feedforward neural networks with the same structure but different parameters are denoted by $FFN_{iu}(\cdot)$, $FFN_{ix}(\cdot)$, $FFN_{uu}(\cdot)$, and $FFN_{ux}(\cdot)$, the following formula gives the computation of student interaction weight, student adaptive weight, music education resource interaction weight, and music education resource adaptive weight $x_{iu}^{(m)}$, $x_{ix}^{(m)}$, $x_{uu}^{(m)}$, $x_{ux}^{(m)}$:

$$\begin{aligned} x_{iu}^{(m)} &= SOFTMAX\left(FFN_{iu}\left(r_i, r_{ivt}^{(m)}\right)\right) \\ x_{ix}^{(m)} &= SOFTMAX\left(FFN_{ix}\left(r_i, r_{ivt}^{(m)}\right)\right) \\ x_{uu}^{(m)} &= SOFTMAX\left(FFN_{uu}\left(r_i, r_{uvt}^{(m)}\right)\right) \\ x_{ux}^{(m)} &= SOFTMAX\left(FFN_{ux}\left(r_i, r_{uvt}^{(m)}\right)\right) \end{aligned} \tag{3}$$

The output of each *FFN* is normalized through the *softmax* function as shown in the following formula:

$$SOFTMAX(a_u) = \frac{\exp(a_u)}{\sum_k \exp(s_k)} \tag{4}$$

The construction logic of the aggregation layer focuses on “feature fusion and semantic enhancement.” Through the synergy of the GNN and the feedforward network, the scattered interaction features are transformed into structured embedding representations in order to strengthen the overall characterization of student interests and resource attractiveness. For the generation of mobile interaction node embedding vectors, the model concatenates the mobile interaction embedding vector containing weighted neighbor features with the original node feature embedding vector so that the student embedding simultaneously includes both “self-learning attributes” and “interest tendencies reflected in interactions,” and the resource embedding simultaneously includes both “self music features” and “attractiveness reflected in interactions.” These concatenated features are processed by a single-layer feedforward neural network and enhanced by the LeakyReLU activation function to increase nonlinear expressive power. The finally generated mobile interaction node embedding vector not only retains the basic attributes of the node but also integrates the dynamic information from interaction relationships. For example, the embedding of a “beginner learner” student will be deeply bound to the interaction features of “beginner practice pieces,” forming a feature representation more in line with the actual learning state.

Another important function of the aggregation layer is to generate adaptive historical mobile interaction feature representations. By integrating the output of the multi-directional attention layer, the model captures the dynamic patterns in interaction sequences, further supporting the recommendation model’s ability to track changes in student interest. The model concatenates the mobile interaction embedding vector with the adaptive embedding vector and generates adaptive historical feature vectors for students and resources through a feedforward neural network. This design enables the feature representation to dynamically respond to changes in the interaction sequence: for students, the vector adjusts in real time according to changes in recent interaction resource types, frequencies, etc., accurately capturing interest migration; for resources, the vector updates along with the characteristics of the interacted student group, reflecting the dynamic changes in their attractiveness. This dynamic feature generation mechanism enables the model to go beyond the limitations of static attributes and truly support the “dynamic adaptive” recommendation need, providing high-value feature input for preference prediction in the subsequent *MLP* layer. Specifically, assume the number of neighbors for students and music education resources is represented by M_i and M_u. The following formulas give the expressions for interaction embedding vectors g_{iu}, g_{uu} and adaptive embedding vectors g_{ix}, g_{ux}:

$$\begin{gathered} g_{iu} = \sum_{m=1}^{M_i} x_{iu}^{(m)} r_{ivt}^{(m)}, g_{ix} = \sum_{m=1}^{M_i} x_{ix}^{(m)} r_{ivt}^{(m)} \\ g_{uu} = \sum_{m=1}^{M_i} x_{uu}^{(m)} r_{ivt}^{(m)}, g_{ux} = \sum_{m=1}^{M_i} x_{ux}^{(m)} r_{uvt}^{(m)} \end{gathered} \tag{5}$$

Assume the connection weights of the feedforward neural networks used to generate student interaction embedding representation and music education resource interaction embedding representation are denoted by Q_i and Q_u, and the biases are denoted by y_i and y_u. The final mobile interaction node embedding representations are:

$$\begin{aligned} g_i &= LeakyRELU\left(Q_i\left[r_i \| g_{iu}\right] + y_i\right) \\ g_u &= LeakyRELU\left(Q_u\left[r_u \| g_{uu}\right] + y_u\right) \end{aligned} \tag{6}$$

The expression for generating the adaptive historical interaction feature representation vectors g'_I and g'_u through feedforward neural networks is given by:

$$\begin{aligned} g'_i &= LeakyRELU\left(Q'_i\left[g_{iu} \| g_{ix}\right] + y'_i\right) \\ g'_u &= LeakyRELU\left(Q'_u\left[g_{uu} \| g_{ux}\right] + y'_u\right) \end{aligned} \tag{7}$$

The core of the MLP layer construction is to transform the feature representations output by the mobile interaction embedding generator into the student's preference probability for the music education resource. Its design directly serves the research objective of "accurately predicting the matching degree between students and resources to achieve personalized recommendation" and is closely based on the modeling logic of student–resource interaction relationships. This layer takes the concatenated vector of the mobile interaction node embedding representations (g_i, g_u) and the adaptive historical mobile interaction feature representations (g_i, g_u) as input. Among them, (g_i, g_u) integrates the inherent attributes of students/resources and the dynamic influences in interaction relationships, while (g_i, g_u) captures the dynamic patterns of interaction sequences. After concatenation, a comprehensive feature vector is formed that contains both static attributes and interaction relationships, as well as dynamic interest and attractiveness changes.

2.3 Learning algorithm

Given a training sample (a, b), the objective of the model is to maximize the predicted probability $\hat{b} = d(a)$ when $b = 1$ and minimize the predicted probability $\hat{b}$ when $b = 0$. Assuming the training set is denoted as F, the binary cross-entropy loss function is defined as:

$$loss = -\frac{1}{v}\sum(b\log d(a) + (1-b)\log(1-d(a))) \tag{8}$$

3 DESIGN OF AN ADAPTIVE LEARNING SYSTEM FOR MUSIC EDUCATION EMPOWERED BY INTERACTIVE MOBILE TECHNOLOGY

The adaptive learning system for music education empowered by interactive mobile technology should be designed with accurately matched resource recommendations as the core driving force and should cover a dynamic response mechanism of "behavior perception – data analysis – network evolution." The proposed system captures learners' micro-level learning behaviors in real time through multimodal interactive interfaces on mobile terminals. These data are instantly transmitted to a multi-directional mobile interaction GAN, which constructs a relational graph of learners' basic proficiency, learning preferences, and resource attributes, and dynamically assigns weights through an attention mechanism. This real-time attention distribution updating mechanism upgrades resource recommendation from "static matching" to "dynamic co-evolution," maintaining synchronization with the learner's current state.

The system adopts a path adjustment mechanism based on learning outcome data. Specifically, when a learner demonstrates difficulty understanding recommended chord exercises, the system does not simply push lower-difficulty content but locates the root cause of the issue through multi-dimensional data. Path adjustment also takes into account the contextual characteristics of mobile learning. The personalized learning goal module decomposes long-term plans into staged sub-goals and dynamically adjusts progress based on the completion quality of daily recommended resources. If the learner fails to complete improvisation-related resources for two consecutive weeks, the system will push a "3-minute improvisation intro" lightweight task via mobile terminal and mark in the goal progress page: "Additional improvisation practice required to meet the standard," while linking previously favorited "simplified accompaniments for pop songs" resources as review reminders. Through the above closed loop of "precise resource pushing–full-scenario behavior tracking–multi-dimensional outcome evaluation–contextual path adaptation–personalized goal calibration," the proposed system leverages the real-time interaction capabilities of mobile terminals to realize truly adaptive evolution in music learning, achieving personalized learning paths for every individual.

4 EXPERIMENTAL RESULTS AND ANALYSIS

From the analysis of Table 1, "Recommendation Effect of Dynamic Interaction Features," it can be seen that on the three datasets GiantMIDI-Piano, KDD Cup 2010 Educational Data, and Mobile Music Interaction Dataset, the model based on dynamic interaction features significantly outperforms the model with static features, with improvement rates of 3.65%, 2.83%, and 1.86%, respectively. This indicates that in the adaptive learning system designed in this paper, the integration of spatiotemporal dynamic features from mobile scenarios and learners' real-time interactive behaviors can effectively enhance the accuracy of resource recommendation. The static feature model relies only on fixed interaction relationships, while the dynamic feature model dynamically tracks learning behaviors and feeds them back into the GAN, dynamically adjusting interaction weights and recommendation strategies, making it more aligned with the dynamic learning scenarios in music education.

Table 1. Recommendation effect of dynamic interaction features

Model	MUSIC-AVQA	GiantMIDI-Piano	KDD Cup 2010 Educational Data	Mobile Music Interaction Dataset
Static Features	0.7569	0.6894	0.7156	0.7156
Dynamic Features	0.7569	0.7126	0.7359	0.7289

Table 2. Recommendation effect of confidence embedding vector

Model	MUSIC-AVQA	GiantMIDI-Piano	KDD Cup 2010 Educational Data	Mobile Music Interaction Dataset
Removing Confidence Embedding Vector	0.7589	0.7123	0.7362	0.7256
Using Positional Encoding	0.7541	0.6895	0.7358	0.7254
Using a Non-trainable Confidence Embedding Vector	0.7598	0.7145	0.7354	0.7248
Using a Randomly Initialized Trainable Vector	0.7456	0.6895	0.7214	0.7261
Confidence Embedding Vector Used in This Paper	0.7523	0.7126	0.7369	0.7282

Table 3. Recommendation performance using different attention functions

Model	MUSIC-AVQA	GiantMIDI-Piano	KDD Cup 2010 Educational Data	Mobile Music Interaction Dataset
Sparse Attention	0.7526	0.7152	0.7326	0.7256
Additive Attention	0.7548	0.7142	0.7355	0.7248
Gated Attention	0.7569	0.7158	0.7324	0.7231
Causal Attention	0.7523	0.7152	0.7316	0.7258
FFN used in this paper	0.7588	0.7162	0.7359	0.7333

The analysis of Table 2, "Recommendation Effect of Confidence Embedding Vector," is as follows: On the three datasets highly relevant to music education—GiantMIDI-Piano, KDD Cup 2010 Educational Data, and Mobile Music Interaction Dataset—the confidence embedding vector model used in this paper outperforms all comparison models. Taking GiantMIDI-Piano as an example, the proposed model achieves a 2.31% improvement over "Using Positional Encoding" and a 2.31% improvement over "Using Randomly Initialized Trainable Vector," and slightly exceeds "Removing Confidence Embedding Vector." In the KDD Cup educational dataset, the proposed model improves by 1.55% over "Using Randomly Initialized Trainable Vector," demonstrating its adaptability to educational scenarios. In the Mobile Music Interaction Dataset, the proposed model shows a 0.26% improvement over "Removing Confidence Embedding Vector," highlighting its optimization effect in mobile interaction scenarios. Although the value is slightly lower on the MUSIC-AVQA dataset, the significant advantages in the other three datasets indicate that the confidence embedding vector designed in this paper better fits the dynamic interaction requirements of music education.

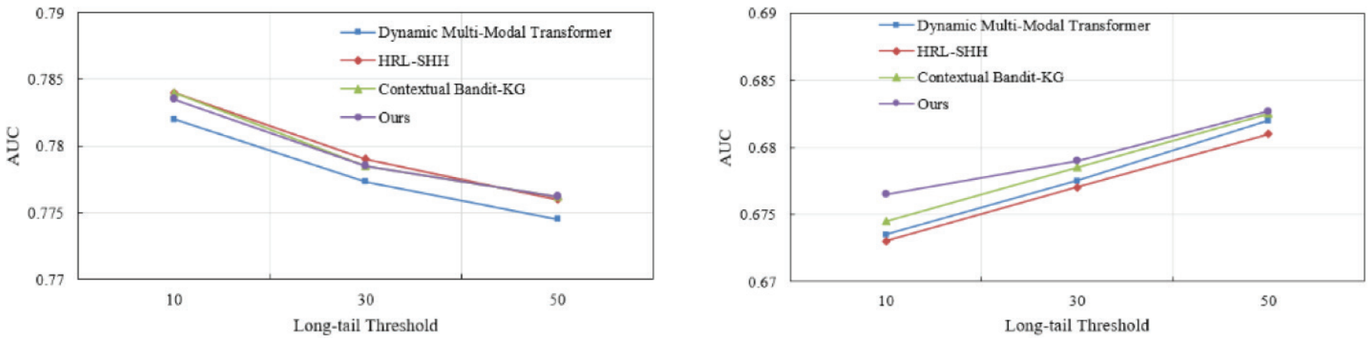
The analysis of Table 3 shows that the FFN attention function used in this paper demonstrates excellent recommendation performance across multiple music education-related datasets, effectively validating the system design. In MUSIC-AVQA, GiantMIDI-Piano, KDD Cup 2010 Educational Data, and Mobile Music Interaction Dataset, the FFN achieves leading or near-leading recommendation scores.

Taking GiantMIDI-Piano as an example, the FFN improves 0.001 over Sparse Attention (0.7152) and 0.0004 over Gated Attention (0.7158), indicating better capture of dynamic interaction relationships in piano learning. In the Mobile Music Interaction Dataset, the FFN score of 0.7333 surpasses Additive Attention's 0.7248 and Gated Attention's 0.7231, confirming its efficient adjustment of interaction weights in fragmented mobile scenarios.

The analysis of Table 4 shows that the model proposed in this paper achieves significantly better AUC performance than most comparison models across the four key datasets, fully validating its effectiveness. Specifically, on MUSIC-AVQA, GiantMIDI-Piano, KDD Cup 2010 Educational Data, and Mobile Music Interaction Dataset, the AUCs of the proposed model are 0.7525, 0.7145, 0.7362, and 0.7258, respectively. Taking the educational scenario dataset KDD-Cup as an example, the proposed model improves by 1.17% compared with the comparison models. In the mobile interaction scenario, it improves by 9.0% over SASRec, reflecting a deep adaptation to core scenarios in music education. From the experimental results, it can be seen that compared with traditional models, the proposed model better aligns with the stepwise progression demands of music learning by integrating the progression pattern of musical skills and dynamic feedback mechanisms. Its significantly better performance on the GiantMIDI-Piano dataset confirms its optimization capability for skill-level hierarchical learning scenarios.

Table 4. AUC results of overall comparative experiments on four datasets

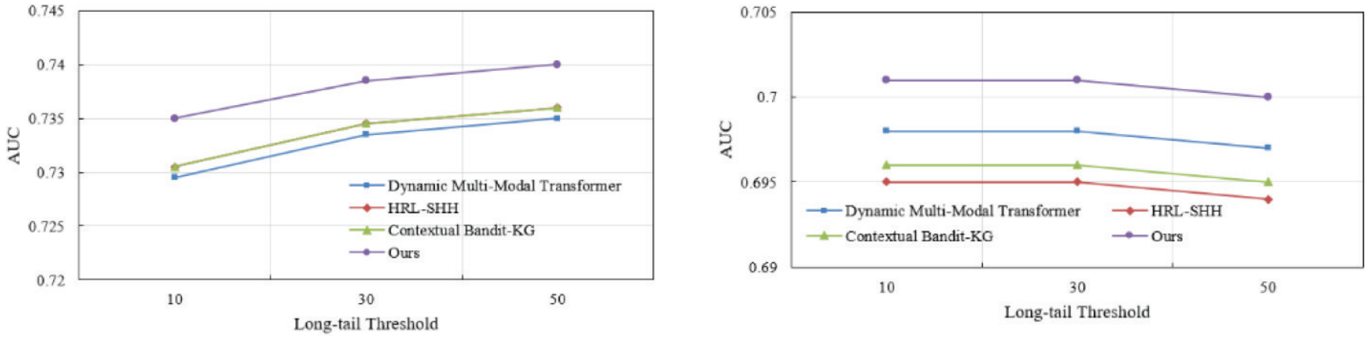
Model	MUSIC-AVQA	GiantMIDI-Piano	KDD Cup 2010 Educational Data	Mobile Music Interaction Dataset
SimCLR	0.6452	0.6123	0.6785	0.6125
SASRec	0.6639	0.6254	0.6894	0.6358
Attention-based RNN	0.7548	0.6885	0.7256	0.6895
Dynamic Multi-Modal Transformer	0.7512	0.7152	0.7245	0.7256
HRL-SHH	0.7569	0.7162	0.7258	0.7245
Contextual Bandit-KG	0.7589	0.7126	0.7236	0.7261
FFN Used in This Paper	0.7525	0.7145	0.7362	0.7258



a) MUSIC-AVQA

b) GiantMIDI-Piano

Fig. 3. *(Continued)*



c) KDD Cup 2010 Educational Data

d) Mobile Music Interaction Dataset

Fig. 3. Recommendation performance comparison under different long-tail thresholds

The analysis of Figure 3 indicates that the proposed model exhibits outstanding recommendation performance under different long-tail thresholds, particularly showing significant advantages in core music education datasets. Taking MUSIC-AVQA (Figure 3a) as an example, as the long-tail threshold increases from 10 to 50, the AUC of the proposed model consistently remains higher than that of the comparison models, maintaining above 0.775 even at threshold 50, reflecting efficient recommendation capability for long-tail resources. In GiantMIDI-Piano (Figure 3b), the AUC of the proposed model shows an upward trend with increasing long-tail thresholds, whereas the comparison models grow slowly or remain flat, validating its improved accuracy in recommending sparse and personalized resources in piano learning, aligning with the demand for niche resources in musical skill progression. In KDD Cup 2010 Educational Data (Figure 3c), the proposed model achieves an AUC of 0.74 at threshold 50, significantly higher than comparison models, proving its precise recommendation ability for long-tail learning resources in educational scenarios. In the Mobile Music Interaction Dataset (Figure 3d), the AUC of the proposed model remains stable above 0.70 across all thresholds, with the smallest fluctuation, demonstrating robust recommendation capability for long-tail resources in mobile scenarios.

5 CONCLUSION

This paper focused on the deep integration of interactive mobile technology and music education and constructed an adaptive learning system based on a multi-directional mobile interaction GAN, realizing dynamic optimization of both resource recommendation and the learning process. At the feature representation level, the system extracted multi-directional interaction and spatiotemporal dynamic features in mobile scenarios, injecting context-awareness into the recommendation model. At the model architecture level, it captured interaction weight differences via a GAN, and dynamically adjusted recommendation strategies in real-time through a mobile interaction-aware module, precisely adapting to the dynamic nature of music learning. At the learning algorithm level, it incorporated the progression rules of musical skill development into the loss function, enhancing feedback optimization for learning outcomes such as performance assessment and music theory testing. Experimental data verified the system's advantages in dynamic interaction, personalized learning path adjustment, and long-tail resource recommendation. Through a closed loop of "resource recommendation – behavior tracking – outcome evaluation – path adjustment – recommendation optimization," the system achieved real-time iteration between learning behavior and recommendation strategies. It provides a technological paradigm for adaptive learning in music education. The innovation lies in the deep coupling of multi-directional mobile interaction and GANs, addressing the core challenges of resource matching and learning path optimization in dynamic scenarios, and carries important theoretical and practical significance.

6 REFERENCES

[1] M. N. Chowdhury, H. M. Z. Haque, K. T. Tahmid, F.-T.-Z. Salma, and N. Ahmed, "A novel approach for product recommendation using smartphone sensor data," *International Journal of Interactive Mobile Technologies (iJIM)*, vol. 16, no. 16, pp. 190–204, 2022. https://doi.org/10.3991/ijim.v16i16.31617

[2] B. Soelistijanto and V. Ayu, "Improving traffic load distribution fairness in mobile social networks," *Algorithms*, vol. 15, no. 7, p. 222, 2022. https://doi.org/10.3390/a15070222

[3] A. M. Almasri, L. B. Gouveia, and S. S. Fernandes, "Estimating the cost of wasted energy during overcharging smartphones and using COVID19 apps," *International Journal of Interactive Mobile Technologies (iJIM)*, vol. 16, no. 14, pp. 201–210, 2022. https://doi.org/10.3991/ijim.v16i14.32021

[4] J. Schampheleer, A. Huss, and M. Deruyck, "Network-scale impact of vegetation loss on coverage and exposure for 5G networks," *IEEE Access*, vol. 13, pp. 23902–23912, 2025. https://doi.org/10.1109/ACCESS.2025.3538054

[5] A. Rahman *et al.*, "Network anomaly detection in 5G networks," *Mathematical Modelling of Engineering Problems*, vol. 9, no. 2, pp. 397–404, 2022. https://doi.org/10.18280/mmep.090213

[6] M. Ouyang, "Employing mobile learning in music education," *Education and Information Technologies*, vol. 28, pp. 5241–5257, 2023. https://doi.org/10.1007/s10639-022-11353-5

[7] M. A. Duarte-García and J. R. Sigal-Sefchovich, "Working with electroacoustic music in rural communities: The use of an interactive music system in the creative process in primary and secondary school education," *Organised Sound*, vol. 24, no. 3, pp. 228–239, 2019. https://doi.org/10.1017/S135577181900030X

[8] M. Rusiñol, J. Chazalon, and K. Diaz-Chito, "Augmented songbook: An augmented reality educational application for raising music awareness," *Multimedia Tools and Applications*, vol. 77, pp. 13773–13798, 2018. https://doi.org/10.1007/s11042-017-4991-4

[9] Y. Jie, "Entertainment robots based on smartphone terminals for mobile network music E-learning mode," *Entertainment Computing*, vol. 52, p. 100787, 2025. https://doi.org/10.1016/j.entcom.2024.100787

[10] Z. Zhang, "Analysis of the application of modern mobile wireless network terminal in the tutoring design of college physical education course teaching," *Wireless Networks*, vol. 30, pp. 5319–5331, 2024. https://doi.org/10.1007/s11276-023-03277-w

[11] J. Zhang, "Research on classroom teaching evaluation and instruction system based on GIS mobile terminal," *Mobile Information Systems*, vol. 2021, no. 1, pp. 1–11, 2021. https://doi.org/10.1155/2021/9790766

[12] B. Zhang, "The sports condition monitoring system based on intelligent mobile terminal," *AGRO Food Industry Hi-Tech*, vol. 28, no. 3, pp. 2661–2664, 2017.

[13] H. Li, "Application of an Internet of things oriented network education platform in English language teaching," *Advances in Multimedia*, vol. 2022, no. 1, pp. 1–10, 2022. https://doi.org/10.1155/2022/9651028

[14] M. Müller, B. McFee, and K. Kinnaird, "Interactive learning of signal processing through music: Making Fourier analysis concrete for students," *IEEE Signal Processing Magazine*, vol. 38, no. 3, pp. 73–84, 2021. https://doi.org/10.1109/MSP.2021.3052181

[15] A. Fanani, K. Hastuti, A. Syarif, and A. Mulyana, "Rule-based interactive learning application model on how to play music instruments," *International Journal of Emerging Technologies in Learning (iJET)*, vol. 15, no. 15, pp. 52–63, 2020. https://doi.org/10.3991/ijet.v15i15.11486

[16] M. K. Shan, F. F. Kuo, M. F. Chiang, and S. Y. Lee, "Emotion-based music recommendation by affinity discovery from film music," *Expert Systems with Applications*, vol. 36, no. 4, pp. 7666–7674, 2009. https://doi.org/10.1016/j.eswa.2008.09.042

[17] J. W. Lee, H. J. Kim, and S. G. Lee, "Conceptual collaborative filtering recommendation: A probabilistic learning approach," *Neurocomputing*, vol. 73, nos. 13–15, pp. 2793–2796, 2010. https://doi.org/10.1016/j.neucom.2010.04.005

[18] Y. Deldjoo, M. Schedl, and P. Knees, "Content-driven music recommendation: Evolution, state of the art, and challenges," *Computer Science Review*, vol. 51, p. 100618, 2024. https://doi.org/10.1016/j.cosrev.2024.100618

[19] A. N. Sharifbaev, H. N. Zainidinov, I. V. Kovalev, I. N. Kravchenko, and Y. A. Kuznetsov, "Increasing the effectiveness of personalized recommender systems based on the integrated GNN-RL model," *Journal of Machinery Manufacture and Reliability,* vol. 53, pp. 980–986, 2024. https://doi.org/10.1134/S1052618824700845

7 AUTHOR

Kai Zhang is Deputy Director of the Department of Music, Cai Yuanpei School of Art and Design, Shaoxing University, Huancheng West Road 508, Shaoxing, China. His main work involves the education of music for the students in the university (E-mail: 18657598885@163.com).