International Journal for Housing Science and Its Applications
Publish August 10, 2025. Volume 46, Issue 4 Pages 2889-2899

https://doi.org/10.70517/ijhsa464240

Innovative Ways of Music Distribution in the Digital Age

Jiang Jiang[1,*]
[1] Xinxiang University Music College, Xinxiang, Henan, 453003, China
Corresponding authors: (e-mail: 17516388187@163.com).

Abstract The continuous development of cloud computing and deep learning technologies has opened up new possibilities for the efficient dissemination of digital music. This paper expands the relevance of query content through data preprocessing and pseudo-correlation feedback techniques. By combining a query likelihood model with a deep ranking learning method based on Pointwise, user preference-based music ranking is achieved. A deep learning-based composite model (ResGRU) is established to extract user implicit behavior data and preferred audio features, enabling precise recommendations for queried music. Research shows that the comprehensive ranking method achieves metric values greater than 0.800 at music popularity levels of 30%, 60%, 90%, and 100%. The conversion rates for both the same network and different networks exceed 87% and 85%, respectively. The ResGRU composite model outperforms five comparison models across six metrics in two datasets, and the best music recommendation results are achieved when the optimal convolutional kernel size is set to 6.

Index Terms music retrieval, digital music dissemination, deep ranking, ResGRU model, music recommendation system

I. Introduction

As a social phenomenon, the dissemination of music has undergone constant change with the development of technology. From oral transmission in primitive societies, to the transmission of musical notation in the agricultural age, to the transmission of records in the industrial age, to the transmission of radio, film, and television in the electronic age, and now to the digital transmission of the digital age, music has continuously presented itself to audiences in new forms under the leadership of technological change [1].

Oral transmission was the most primitive form of music dissemination, but due to its poor recordability, it could not be effectively preserved. The emergence of musical notation made the cross-temporal and spatial dissemination of musical art possible, but due to the lowpopularity of musical notation, music could only become the private property of some professionals and a few upper-class individuals [2]. In 1877, Edison demonstrated the phonograph to the world, enabling music to be formally recorded [3]. The advent of the record era gave music dissemination a mass-market character, with records becoming the first medium for music dissemination. Record production severed the visual presentation of musical works, retaining only the temporal attributes of sound. Audiences became 'blind' music listeners, with the way of appreciating music shifting from a combination of sight and sound to listening solely with the ears, reducing visual engagement. Meanwhile, the spatial scope of music dissemination expanded, but the temporal scope of dissemination contracted due to technological intervention [4]-[8].

In the digital age, music is ubiquitous through various forms of media, including in-car radio, music videos, and film and television soundtracks. The birth and development of the internet in the 1990s ushered us into a new, high-speed information age. Daniel [9] noted that during the same period, CDs dominated the music industry's distribution format, marking the beginning of the industry's digital transformation. In 1993, the emergence of MP3 audio compression technology enabled music to be infinitely replicated and distributed, breaking free from traditional one-way broadcast-style transmission. Point-to-point interactive transmission became a reality [10], [11]. Subsequently, the widespread adoption of computers, mobile phones, and smart devices made music distribution more convenient. Related music streaming platforms were developed, allowing users to freely download, distribute, and exchange music according to their preferences on the internet or these platforms. Music culture dissemination also exhibited characteristics of massive, interactive distribution [12], [13].

With the continuous development of high-speed internet technology and information network technology, the music market in the digital age has undergone significant changes, making the acquisition of music resources more convenient and efficient, such as through streaming platforms. Guo [14] noted that after the emergence of simple digital music formats, the rise of streaming services (such as Apple Music) has reshaped the ways in which music is acquired and monetised. Music streaming platforms leverage the internet to facilitate the dissemination and promotion of music, enabling music works to be freely shared on a global scale [15]. Guo [16] reviewed the

dissemination and evolution of popular music culture, which has evolved through individual performances, bands and music groups, digital media, and internet-based dissemination. The review particularly highlighted innovative promotional strategies for popular music on social media, where music discovery and promotion exhibit viral trend phenomena. Taking Douyin as an example, creators on the platform add music to their videos or images, with some content focusing on sharing music, concerts, playlists, etc. In 2024, the total number of music plays reached 3.2 trillion, and the platform also launched the 'Douyin Sees Music' initiative, with participating songs generating 118.6 billion plays. Lyu [17] analysed the visual dissemination of ethnic music on the Douyin short video platform. These videos combine audio and video elements, providing a more interactive dissemination model for music.

In the digital age, music-related technologies are constantly emerging. O'Kane [18] introduced immersive audio technology, including Dolby Atmos, which provides listeners with a three-dimensional spatial perception and an immersive experience when listening to music. Dolby Laboratories has consistently worked to improve surround sound technology, and Dolby Atmos not only transforms the way music is listened to but also influences creative styles. As a result, the Dolby Atmos music libraries on major music streaming platforms have been steadily expanding in recent years.

In addition, artificial intelligence technology is reshaping the music industry, encompassing creation, distribution, and consumption [19]. Mokoena and Obagbuwa [20] discussed the application of AI technology in digital music streaming platforms, where AI automation technology can be used to achieve personalised music recommendations and improve subscription services. Music recommended through algorithmic personalisation has effectively facilitated the dissemination of music. Cao [21] created a central station and branch centre system tailored for the dissemination of ethnic music to preserve the traditional environment of minority ethnic music, enabling ethnic music to be better disseminated in the context of globalisation. Additionally, Tencent Music Entertainment Group launched China's first virtual music platform, TMELAND, which attracted 12 million viewers for its virtual concerts. In a survey study by Onderdijk et al. [22] on virtual reality concerts, it was found that the share of virtual reality concerts in the music market is steadily increasing, and 70% of viewers believe this will become the future of the music industry.

This paper utilizes cloud storage and cloud computing technologies to enhance the management, search efficiency, and portability of digital music resources. It standardizes the format of searchable recommendation information, establishes a multi-dimensional information index, and performs standardized preprocessing of search queries and music data. Using pseudo-correlation feedback technology, it extracts information expansion terms and combines them with a query likelihood model for initial ranking and weight calculation of search results. It then integrates a pointwise deep sorting learning method to achieve the optimal sorting results. It converts users' implicit data into music ratings and introduces Mel frequency scale (MEL) filtering to map users' preferred specific audio content. By comprehensively processing user data and audio data using the ResGRU model, it achieves precise music query recommendations.

II. Digital music management and distribution based on cloud computing and deep learning

II. A.Application of Cloud Computing in Music Distribution

The core concept of "cloud computing" is to centrally manage and schedule a large number of network-connected computing resources, forming a computing resource pool that provides services to users on demand. In the music industry, this manifests as: integrating massive digital music resources, centrally storing them, and using efficient algorithms, management and scheduling software, and hardware resources to promptly and quickly respond to user needs, distributing music resources to different users. For end-users, there is no need to carry any audio or video media. As long as there is an internet connection and a device, users can access the desired resources from the cloud at any time.

The rapid development of digital music has led to an abundance of music resources. Music enthusiasts often have personal music libraries containing tens or even hundreds of gigabytes of music works, but the storage, copying, and retrieval of these resources have gradually become their biggest challenges, hindering the further development of digital music dissemination. The emergence of "cloud" technology is aimed at addressing this issue, with music cloud computing driving greater intelligence in mobile music applications. Cloud computing can be simply represented as cloud storage and cloud computing. Through cloud storage, users can create their own private cloud with a cloud music provider. Combined with relevant music software, music content stored in the cloud can be played and shared on various devices such as mobile phones, computers, and televisions, eliminating the need for users to spend time and effort copying music from their computer storage to other devices. This fully leverages the interactive nature of the internet. Additionally, through cloud computing, users can quickly and conveniently search for music works that meet their requirements. Users only need to be aware of the cloud's existence and do not need to understand its inner workings. Currently, factors such as broadband limitations in certain regions, high data

charges, and copyright issues have somewhat constrained the development of cloud music. It is foreseeable that, under the high priority given by the national government, with the continuous emergence of new technologies and the ongoing optimization of music-sharing models, issues such as broadband, data charges, and copyright will inevitably be resolved. At that point, cloud music will usher in a bright future.

II. B.Deep learning methods and technical implementation

II. B. 1) Data preprocessing

During the process of users querying online music information, the system ranks the output results based on the keywords and their frequency in the online music information, as well as the query content entered by the user. Before the search program is initiated, both the query data and music data must undergo preprocessing, which involves creating information-rich indexes to enhance the quality of the query results. Some of the music in the network is in digital format, so it needs to be converted into a unified text format. This paper establishes four types of indexes based on the specific content of the text information: overall data index, metadata (keywords provided by the author) index, content (summary of music content) index, and social information (user comments and tags assigned to them) index.

II. B. 2) Pseudo-correlation feedback technology

To obtain more query results, you can select expanded words with meanings similar to the keywords based on the user's initial query content, combine them with the input content to form expanded query phrases, and thereby improve the richness, accuracy, and completeness of the query content. Pseudo-related feedback technology can effectively achieve query expansion, and the expanded terms obtained through this technology represent the maximum expansion of the query content. After obtaining the first batch of online music information based on the user's original query, expanded terms are extracted from the top-ranked information to further enrich the query content. It is evident that the quantity and accuracy of the expanded terms are determined by this portion of online music information.

The pseudo-correlation feedback technique is based on the assumption that the top $1 \sim k$ ranked online music information obtained from the query results based on the user's original query content are indeed relevant to the keywords entered by the user. The process for initial sorting of query results based on this assumption is as follows.

Step 1: Perform an initial sorting of the query results based on the original content entered by the user, and select the top k ranked online music information based on the strength of content relevance.

Step 2: Extract content keywords from the above k pieces of online music information, and use the top w most frequently occurring words as expanded keywords related to the user's input.

Step 3: Perform a secondary query using the new query phrases formed by combining the user's input keywords with the expanded keywords, and obtain new query results.

Step 4 Use the search engine's query likelihood model to perform an initial sorting of the above results, with the ranking based on relevance calculated using the following formula:

$$\log P\left(Q|D\right) = \sum_{i=1}^{m} \log \frac{f_{q_i,D} + \mu \frac{c_{q_i}}{|C|}}{|D| + \mu} \tag{1}$$

In the equation, $f_{q_i,D}$ represents the word frequency of keyword q_i in online music information; $\frac{c_{q_i}}{|C|}$ represents the distribution probability of keyword q_i in the dataset; $|D|$ is the similarity of the initially sorted music; μ is the Dirichlet smoothing parameter.

II. B. 3) Reordering Mechanism

For the retrieval system designed in this paper, in order to expand the scope of the query target, it is first assumed that music similar to the user's preferred music type can also be regarded as music preferred by the user. Based on this assumption, the reordering can be performed using the following formula:

$$score_r\left(Q,D\right) = a \times score\left(Q,D\right) + \left(1-a\right)e\left(D\right) \tag{2}$$

In the formula, $score\left(Q,D\right)$ represents the initial ranking score; $e(D)$ represents the degree to which similar music plays a role, and its value is calculated using similarity weighting, i.e.:

$$e(D_i)=\sum_{\substack{D_j\in D\\ D_j\neq D_i}} sim(D_i,D_j)\times score(Q,D_j) \tag{3}$$

In the formula, $sim(D_i,D_j)$ represents the similarity between the i th and j th songs in the initial ranking results. Q is the set of songs.

If the number of occurrences of a single domain in the online music data is used as the basis for reordering, then the similarity between two songs is calculated as follows:

$$Sim(D_i,D_j)=\cos\langle vec_{Di},vec_{Dj}\rangle \tag{4}$$

In the formula, vec_{D_i} represents the feature vector of the music ranked i; vec_{D_j} represents the feature vector of the music ranked j. If the similarity of the combined features is used as the basis for reordering, then the similarity calculation method based on the initial sorting results is

$$sim_1(D_i,D_j)=\begin{cases}1 & D_j\in S(D_i)\,|\,D_i\in S(D_j)\\ 0 & D_j\notin S(D_i)\,\&\,D_i\notin S(D_j)\end{cases} \tag{5}$$

In the equation, $S(D_i)$ represents music similar to the type of D_i. For the calculation results of equation (5), if the music in $S(D_i)$ is included in the initial sorting results, then the tag value of that music is 1.0; otherwise, it is 0.0. Thus, the re-sorting results can be obtained by comparing the similarity. Based on this, the re-ranking mechanism established in this paper is DTN, ITN, DT, IT, TN, T, N, I, and D, where T, N, I, and D represent user tags, browsed nodes, similar music, and similar music of similar music, respectively.

The sorting results obtained under the above-mentioned reordering mechanisms need to undergo an organic integration process to ensure sorting accuracy. However, current sorting integration methods mostly rely on manual or semi-manual approaches, which can easily lead to biases in the final results. Therefore, this paper adopts a pointwise deep sorting learning method to obtain optimal sorting results. Pointwise is primarily used for processing individual documents. It can convert documents into feature vectors in a specific manner, enabling sorting through regression or classification in a machine learning framework. During the sorting learning process, multiple different sorting results for music are used as feature vectors, with the ranking of music in the optimal sorting results serving as the training objective. The mapping relationship between multiple sorting results and the optimal ranking is obtained, serving as the basis for selecting the best reordering mechanism.

II. C.Music Recommendation System Model

II. C. 1) User-song rating matrix construction

Music recommendation systems typically collect users' implicit behavioral data, such as browsing history, playback records, and collection behavior, which contrasts with movie recommendation systems that rely on users' explicit behavioral data (e.g., ratings). Although the information contained in implicit feedback may not be as direct and clear as explicit ratings, the sheer volume of data can still reveal users' true preferences. Additionally, considering that many users often do not participate in rating activities on the platform, analyzing user preferences using implicit feedback data becomes an inevitable choice. This method not only compensates for the limitations of explicit data but also captures more rich and in-depth user preference information through detailed user behavior analysis. Therefore, in the absence of direct evaluations, the widespread collection and analysis of implicit data becomes particularly critical. By recording and evaluating users' behavioral trajectories, it provides music recommendation systems with a more nuanced and comprehensive understanding of user preferences. This method not only elegantly circumvents the issue of low user engagement but also ensures the personalization and accuracy of recommendation results through its large sample size and rich behavioral dimensions. In summary, by carefully analyzing implicit feedback data, it is possible to uncover subtle differences in user preferences, thereby providing music recommendation services that better align with user needs, enhancing their experience on digital music platforms. Most users do not directly express their level of liking for music, but their behavior—playback frequency, listening duration, song collections, etc.—provides rich indirect evidence to reveal their true preferences. Therefore, by mining and analyzing these detailed behavioral data, a more subtle and precise recommendation mechanism can be established to adapt to users' ever-changing music tastes and listening scenarios. The method proposed in the References section of this paper converts the implicit data of the number of times a user plays a song into a score for the implicit evaluation of the music. Combined with advanced recommendation algorithms, this method

not only enhances the user experience, but also guides users through the sea of music recommendations, helping them discover and explore more songs that they may not have encountered but may potentially like. To this end, this paper uses the implicit data of users' song playback counts as the core data source. By analyzing the frequency of song playback, it calculates users' implicit ratings for each track. Specifically, based on users' historical listening records, this paper counts the playback frequency of each song in the records. Then, using the algorithm proposed in this paper, these playback frequencies are converted into specific scores, thereby constructing a comprehensive scoring matrix that reflects users' music preferences. The specific algorithm description is as follows:

1) Nonlinear Transformation

To obtain the number of times each song has been played by users, we first apply the logarithmic formula (as shown in Equation (6)) to convert the number of times each song has been played, thereby reducing the gap between the maximum and minimum values of the number of times songs have been played by users. x is the original number of times a song has been played by a user, and y is the converted value of the number of times a song has been played by a user:

$$y = \log(1.0 + x) \tag{6}$$

2) Standardization processing

Next, the number of plays for each user song after logarithmic conversion is standardized so that the user's song rating scores are distributed between 1.0 and 10.0. Standardization processing can be achieved through maximum-minimum standardization using the following formula (7):

$$z = 1.0 + \left(9.0 \times \frac{y - \min(Y)}{\max(Y) - \min(Y)} \right) \tag{7}$$

In this context, Y represents the set of play counts for each song in the user's song playback history, where each play count has been transformed using the aforementioned logarithmic transformation. $\min(Y)$ denotes the minimum value within the set, $\max(Y)$ denotes the maximum value within the set, and z represents the user's final rating for the songs they have listened to.

Through the above two steps, this paper effectively converts the user's original playback counts for songs into a more reasonable rating range. This paper sets the maximum rating for songs to 10.0, reducing the impact of the long-tail music distribution on ratings. After processing through the above two methods, even songs with fewer playback counts can receive fair evaluations while maintaining sensitivity to user preferences. By applying the above two-step process to all users' historical listening records, the final user-song rating matrix can be obtained.

II. C. 2) Audio Feature Extraction

To classify users' preferred music genres, it is necessary to analyze audio signals. However, these audio files are typically stored in formats such as WAV or MP3, which only contain waveform data and cannot be directly used for training by deep learning models. Therefore, it is essential to first convert the audio signals into two-dimensional spectrograms to enable effective feature extraction using deep learning models. In fact, audio features provide the most effective and distinctive information for distinguishing between different songs. Therefore, extracting audio features from songs is a critical step in this study. Audio features can be broadly categorized into two types: time-domain features and frequency-domain features. To comprehensively extract and represent audio features, this paper proposes a feature representation method that can comprehensively reflect both the time-domain and frequency-domain information of audio.

Compared to waveform diagrams, time-frequency spectra provide richer frequency domain information. However, this information represents the full frequency range of sound. Although the human ear can perceive sound frequencies ranging from 20.0 Hz to 20,000.0 Hz, human sensitivity to high-frequency sounds decreases as frequency increases, meaning that the ability to perceive changes in frequency diminishes as frequency rises. This means that at higher frequencies, the human ear becomes less sensitive to changes in sound. For example, in an audio sample where the frequency increases from 1000.0 Hz to 2000.0 Hz at two different time points, despite the frequency doubling, human perception of pitch does not increase linearly in proportion. This indicates that there is a nonlinear relationship between human perception of pitch changes and the actual increase in frequency. To address this issue, the Mel frequency scale (MEL) was introduced, which converts ordinary Hertz frequency f into Mel frequency $(mel(f))$ through a mapping formula. This conversion relationship is expressed in Formula (8):

$$mel(f) = 2595.0 \times \log_{10}\left(1.0 + \frac{f}{700}\right) \tag{8}$$

Through the above Mel mapping formula, the Mel frequency obtained is linearly related to human auditory perception. In other words, when the Mel frequency doubles, the pitch perceived by the human ear also doubles accordingly. Given that humans have different sensitivities to sounds of varying frequencies, this mapping acts like a filter, selectively perceiving sounds at specific frequencies. Building on this, some researchers have employed Mel filter banks, using a series of filters to map the time-frequency spectrum of an audio signal into a Mel spectrum diagram.

II. C. 3) ResGRU Model

This paper combines the advantages of ResNet and GRU to propose a composite model that integrates ResNet and GRU, called the ResGRU model. The model aims to fully utilize the advantages of ResNet in feature extraction while leveraging the advantages of GRU in processing time-series data to achieve effective extraction and utilization of spatiotemporal features. Figure 1 shows the overall structure of the ResGRU model. The model mainly consists of three parts: spatial feature extraction, temporal feature extraction, and output.

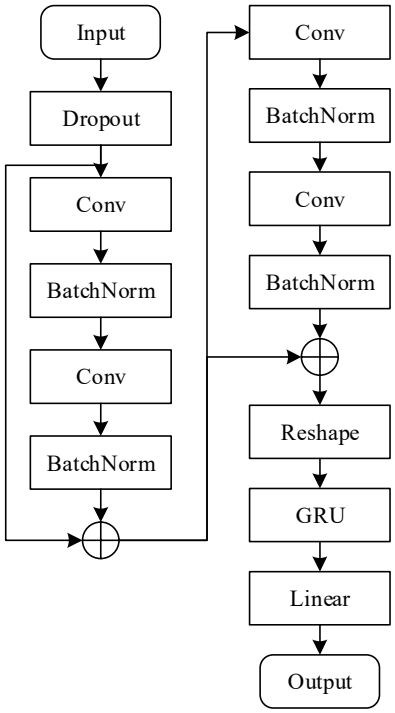

Figure 1: ResGRU Model

The spatial feature extraction component employs a ResNet architecture. The input data first passes through a Dropout layer to prevent overfitting. Subsequently, the data undergoes feature extraction through two residual modules in sequence. Each residual module consists of two convolutional units, whose structure mirrors that of a standard residual block, comprising two convolutional layers and a connection layer. Through the residual modules, high-level spatial features of the input data are extracted while mitigating the vanishing gradient problem inherent in deep networks.

The temporal feature extraction part uses a GRU structure. GRU is an improved recurrent neural network that introduces update gates and reset gates compared to standard RNNs, enabling it to better capture long-range dependencies while alleviating the gradient vanishing and gradient explosion problems. The spatial features extracted by ResNet first pass through a Reshape layer to convert the feature maps into a shape suitable for GRU input. Then, the feature sequence is fed into a GRU layer, which models the temporal dependencies of the feature sequence and extracts temporal features.

The output layer uses a fully connected layer to map the temporal features extracted by the GRU to the output space. Finally, the final prediction result is obtained via the Softmax function.

The forward propagation process of the ResGRU model can be represented by the following formula:

$$X_{dropout} = Dropout(X) \tag{9}$$

$$X_{res1} = \mathrm{Re}\,sBlock1\left(X_{dropout}\right) \tag{10}$$

$$X_{res2} = \mathrm{Re}\,sBlock2\left(X_{res1}\right) \tag{11}$$

$$X_{reshape} = \mathrm{Re}\,shape\left(X_{res2}\right) \tag{12}$$

$$H_{gru} = GRU\left(X_{reshape}\right) \tag{13}$$

$$Y_{pred} = Soft\max\left(Dense\left(H_{gru}\right)\right) \tag{14}$$

In this context, X denotes the input data, $X_{dropout}$ denotes the data processed by the Dropout layer, X_{res1} and X_{res2} denote the outputs of the two residual modules, $X_{reshape}$ denotes the feature sequence processed by the Reshape layer, H_{gru} denotes the output of the GRU layer, and Y_{pred} denotes the final prediction result.

The ResGRU model fully leverages the advantages of ResNet and GRU, enabling it to simultaneously extract spatial and temporal features from the input data. By utilizing residual connection technology, the model can construct a deeper network structure, thereby enhancing its feature extraction capabilities. Additionally, the GRU layer effectively models temporal dependencies, capturing long-range dependencies in sequence data.

III. Effectiveness evaluation of a digital music system based on cloud computing and deep learning

III. A. Verification of audio feature extraction effectiveness

Extracting and processing specific audio features that users prefer is a critical step in achieving precise music recommendation through queries. To validate the effectiveness of the feature representation method selected in the music recommendation model constructed in this paper, two different audio feature extraction methods were chosen as comparison methods. By comparing the results of audio feature extraction experiments, the effectiveness of the feature representation method proposed in this paper was verified. Figure 2 shows the comparison of audio feature extraction results among the three methods. In the fundamental frequency period of 0.00–0.05 seconds, the three feature extraction methods were applied to analyze 300 frames of audio. It was found that compared to the autocorrelation feature extraction method and the average amplitude difference feature extraction method, the feature representation method extracted audio features with more fluctuations, indicating that the extracted audio features were richer and more detailed. Therefore, it can be concluded that the feature representation method has better audio processing performance, enabling the identification and detection of characteristics in users' preferred music, thereby aiding in the precise recommendation of music for queries.

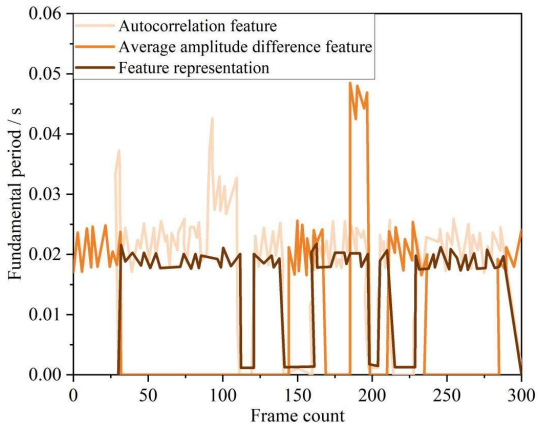

Figure 2: Comparison of audio feature extraction effects among three methods

III. B. Query Sorting Effect Verification

III. B. 1) Comparison of performance differences between different sorting methods

Four query content sorting methods, RankNet, ListNet, RankSVM, and LambdaMART, are selected as the comparison methods, and the sorting performance is compared with the comprehensive sorting method in this paper on different datasets. The correlation (NDCG@15) and mean mean accuracy (MAP) of the top 15 ranking results were used as evaluation indicators. Table 1 shows the NDCG@15 comparison results of different sorting

methods on digital music datasets with different popularity. Table 2 shows the MAP comparison results of different sorting methods on digital music datasets with different popularity. At 30%, 60%, 90% and 100% of the music popularity, the NDCG@15 and MAP values of the ranking method in this paper exceeded 0.800, and were higher than those of the comparison method. They are: 0.871, 0.894, 0.904, 0.966 (NDCG@15); 0.803, 0.890, 0.912, 0.973(MAP). The sorting relevance and average accuracy of the proposed method are better than the comparison method, and can better sort digital music according to the query content entered by the user.

Table 1: NDCG@15 of different sorting methods on different music datasets

	iMusic/$H_{lp=30\%}$	iMusic/$H_{lp=60\%}$	iMusic/$H_{lp=90\%}$	iMusic/$H_{lp=100\%}$
RankNet	0.395	0.410	0.453	0.502
ListNet	0.402	0.505	0.526	0.573
RankSVM	0.537	0.598	0.627	0.672
LambdaMART	0.694	0.702	0.781	0.835
Our method	0.871	0.894	0.904	0.966

Table 2: MAP of different sorting methods on different music datasets

	iMusic/$H_{lp=30\%}$	iMusic/$H_{lp=60\%}$	iMusic/$H_{lp=90\%}$	iMusic/$H_{lp=100\%}$
RankNet	0.371	0.408	0.462	0.521
ListNet	0.424	0.513	0.578	0.580
RankSVM	0.550	0.576	0.636	0.685
LambdaMART	0.619	0.721	0.795	0.841
Our method	0.803	0.890	0.912	0.973

III. B. 2) Comparison of music ranking results for different user groups

As revealed by the comparative experiments in Section 3.2.1, this paper employs a combination of query likelihood models and pointwise deep ranking learning methods for query content ranking, achieving the best ranking performance on the iMusic dataset. Furthermore, this paper utilizes a music product provided by a certain company as an actual application scenario, integrating the proposed method into a music recommendation system via an interface to address music ranking issues, thereby validating the practical application value of the proposed method. This paper first divides the user base of a certain music product into on-network users and off-network users. On-network users refer to users who use network services provided by telecommunications operators, while off-network users refer to users who do not use network services provided by telecommunications operators. Due to the significant differences in natural attributes and social class characteristics between the two user groups, this paper conducts separate experiments for the two user groups. For both the on-network user group and the off-network user group, the comprehensive ranking method is compared via A/B testing to assess conversion rates with and without integration. The conversion rate for the on-network user group is referred to as the on-network conversion rate, while the conversion rate for the off-network user group is referred to as the off-network conversion rate. Figure 3 shows the on-network conversion rate and off-network conversion rate of the comprehensive method under connected and unconnected conditions. After applying the method, the on-network conversion rate exceeded 87% and exhibited more stable fluctuations within a daily timeframe. Similarly, the off-network conversion rate exceeded 85% and also showed more stable fluctuations within a daily timeframe. Compared to the highest on-network conversion rate of 81.91% and off-network conversion rate of 81.23% when not connected, the user conversion rate is higher after applying the sorting method, and the conversion efficiency is more stable. This fully demonstrates that stable and efficient query sorting can keep users engaged in the music recommendation system they use, and users are more satisfied with the search sorting method.

III. C. Music Recommendation Effectiveness Evaluation

III. C. 1) Performance Comparison of Different Recommendation Models

The proposed ResGRU composite model is compared with other benchmark models to evaluate its performance in music recommendation tasks. Table 3 shows the comparison results of recommendation performance across different methods on the KuaiRec dataset. Table 4 shows the comparison results of recommendation performance across different methods on the ML-1M dataset. In terms of performance evaluation metrics such as Recall@25, Recall@60, Recall@100, Mrr@25, Mrr@60, and Mrr@100, the ResGRU composite model outperforms the other five comparison models in both the KuaiRec and ML-1M datasets. On the KuaiRec dataset, the six metric values of the ResGRU composite model are 0.4704, 0.5735, 0.7581, 0.0995, 0.1231, and 0.1459; On the ML-1M dataset, the corresponding metric values are 0.5639, 0.6029, 0.8784, 0.1103, 0.1524, and 0.1887. This indicates that under

consistent model parameter conditions, the ResGRU composite model demonstrates superior music query recommendation performance.

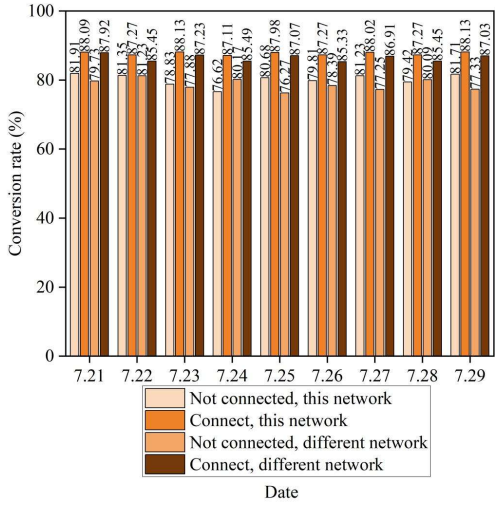

Figure 3: Conversion rate of the method when connected and disconnected

Table 3: Recommendation performance of different methods on KuaiRec dataset

Model	Recall@25	Recall@60	Recall@100	Mrr@25	Mrr@60	Mrr@100
FPMC	0.3242	0.4102	0.5117	0.0736	0.0287	0.0798
GRU4Rec	0.2284	0.4624	0.5652	0.0927	0.1064	0.0992
Caser	0.3223	0.3994	0.4974	0.0281	0.0594	0.0868
SASRec	0.2823	0.4598	0.5661	0.0188	0.1017	0.0946
SINE	0.3722	0.3995	0.4974	0.0722	0.0856	0.0784
ResGRU	0.4704	0.5735	0.7581	0.0995	0.1231	0.1459

Table 4: Recommendation performance of different methods on ML-1M dataset

Model	Recall@25	Recall@60	Recall@100	Mrr@25	Mrr@60	Mrr@100
FPMC	0.4177	0.4396	0.6302	0.0844	0.0428	0.0926
GRU4Rec	0.3219	0.4918	0.6855	0.1035	0.1257	0.1122
Caser	0.4158	0.4288	0.6177	0.0389	0.0787	0.0996
SASRec	0.3758	0.4892	0.6864	0.0296	0.1231	0.1074
SINE	0.4657	0.4289	0.6177	0.0483	0.1049	0.0912
ResGRU	0.5639	0.6029	0.8784	0.1103	0.1524	0.1887

III. C. 2) Analysis of the optimal convolution kernel size

Given the superior recommendation performance of the ResGRU composite model, this study further investigates the impact of its convolutional kernel on the model's deep learning capabilities to identify the optimal kernel size, adjust model parameters, and achieve the best music recommendation results. In this section's experiments, other hyperparameters are kept at their optimal values, with only the convolutional kernel size k adjusted, selecting appropriate values from {3, 6, 9, 12, 15}. Figure 4 shows the impact of adjusting the convolutional kernel size on the results of top-25 recommendations across two different datasets. When the convolutional kernel size k is 6, the ResGRU composite model achieves the highest Recall and Mrr metric values in the top-25 recommendations for both datasets, namely 0.35, 0.34, 0.36, and 0.35. Therefore, the optimal convolutional kernel size for the ResGRU composite model is set to 6 layers to achieve the highest recommendation accuracy.

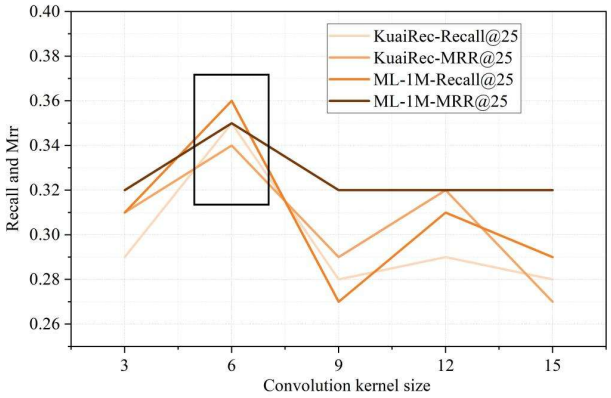

Figure 4: Impact of convolution kernel adjustment on recommendation performance

IV. Conclusion

This paper integrates multiple technologies to construct a digital music recommendation system to optimize digital music query and improve recommendation accuracy. The feature representation method can extract more subtle audio features in 300 frames of gene cycle audio of 0.00-0.05s. The NDCG@15 of the four heat values of the comprehensive sorting method reached 0.871, 0.894, 0.904 and 0.966, and the MAP reached 0.803, 0.890, 0.912 and 0.973. The conversion rate of local/cross-network after accessing the comprehensive sorting method is greater than 87%/85%, which is higher than that of 81.91%/81.23% when not connected. In the comparison of the recommendation accuracy of the six music recommendation models, the evaluation indexes of the ResGRU composite model are higher than those of the comparison model. When the convolutional kernel size is set to 6, ResGRU's top-25 recommendation accuracy is the highest. In view of the scale of digital music, in order to improve the efficiency of sorting and recommendation, the introduction of lightweight modules can be further explored in the future to achieve efficient operation of the system.

References

[1] Hopkins, P. (2017). Ways of transmitting music. In The Garland Encyclopedia of World Music (pp. 90-111). Routledge.
[2] Seebass, T. (2017). Notation and Transmission in European Music History. In The Garland Encyclopedia of World Music (pp. 49-57). Routledge.
[3] Pierce, J. M. (2017). Writing at the Speed of Sound: Music Stenography and Recording beyond the Phonograph. 19th-Century Music, 41(2), 121-150.
[4] Maalsen, S., & McLean, J. (2018). Record collections as musical archives: Gender, record collecting, and whose music is heard. Journal of Material Culture, 23(1), 39-57.
[5] Ceribašić, N. (2021). Music as Recording, Music in Culture, and the Study of Early Recording Industry in Ethnomusicology. International review of the aesthetics and sociology of music, 52(2), 323-354.
[6] Marshall, L. (2019). Do people value recorded music?. Cultural Sociology, 13(2), 141-158.
[7] Whitehouse, S. (2023). “Taking a chance on a record”: lost vinyl consumption practices in the age of music streaming. Consumption Markets & Culture, 26(1), 64-80.
[8] Kiresci, A. (2023). The impact of innovative technologies on small players in the recorded music sector: A chronological overview. Creative Industries Journal, 16(1), 96-111.
[9] Daniel, R. (2019). Digital disruption in the music industry: The case of the compact disc. Creative Industries Journal, 12(2), 159-166.
[10] Cunningham, S., & McGregor, I. (2019). Subjective evaluation of music compressed with the ACER codec compared to AAC, MP3, and uncompressed PCM. International Journal of Digital Multimedia Broadcasting, 2019(1), 8265301.
[11] Boothby, H. (2021). Music for Universities: Composing with MP3 and iPod. Artifact & Apparatus: Journal of Media Archaeology, 1, 1-21.
[12] Caixinha, H., & Caixinha, S. (2019). Connecting the worlds of music with a digital repository for collaborative research, education and content dissemination. In EDULEARN19 Proceedings (pp. 1837-1847). IATED.
[13] Krogh, M. (2019). Music-Genre Abstraction, Dissemination and Trajectories. Dansk Musikforskning Online, 9, 83-108.
[14] Guo, X. (2023). The evolution of the music industry in the digital age: From records to streaming. Journal of Sociology and Ethnology, 5(10), 7-12.
[15] Qu, S., Hesmondhalgh, D., & Xiao, J. (2023). Music streaming platforms and self-releasing musicians: the case of China. Information, communication & society, 26(4), 699-715.
[16] Guo, X. (2023). From solo performances to social media: dissemination and evolution of pop music culture. Media and Communication Research, 4(8), 66-71.
[17] Lyu, Y. (2022, December). Visualization Communication of Ethnic Music on Short Video Platforms A Case Study of the Mongolian Music on Douyin. In 2022 5th International Conference on Humanities Education and Social Sciences (ICHESS 2022) (pp. 363-373). Atlantis Press.
[18] O'Kane, J. (2024). Immersive audio is changing the tune of the music industry; More studios are embracing the sound technology, with Dolby Atmos leading the way. Globe & Mail (Toronto, Canada), R3-R3.

[19] Seneadza, J. S., Boateng, S. L., Marfo, J. S., Boateng, R., & Budu, J. (2025). Transformative Impacts of Artificial Intelligence on the Music Industry: A Narrative Review. AI and the Music Industry, 34-58.
[20] Mokoena, N., & Obagbuwa, I. C. (2025). An analysis of artificial intelligence automation in digital music streaming platforms for improving consumer subscription responses: a review. Frontiers in Artificial Intelligence, 7, 1515716.
[21] Cao, K. (2022, June). Global Dissemination of Minority Music Under the Background of the Internet: based on the Technology of Multimedia Central Station. In 2022 7th International Conference on Communication and Electronics Systems (ICCES) (pp. 891-894). IEEE.
[22] Onderdijk, K. E., Bouckaert, L., Van Dyck, E., & Maes, P. J. (2023). Concert experiences in virtual reality environments. Virtual Reality, 27(3), 2383-2396.