Proceedings of the 2nd International Conference on Machine Learning and Automation
DOI: 10.54254/2755-2721/109/20241353

Deep Learning Based Music Recommendation Systems: A Review of Algorithms and Techniques

Yixiang Liu

Faculty of Arts, University of Nottingham, Nottingham, United Kingdom

hvyyl8@nottingham.ac.uk

Abstract. This paper provides a thorough examination of the utilisation of deep learning in music recommendation systems, which have transformed consumer discovery and engagement with music on streaming platforms. Scalability challenges and the cold-start problem are among the constraints that conventional recommendation methods, such as content-based filtering and collaborative filtering, encounter, which hinder their ability to deliver personalised recommendations that are effective. The processing of multi-modal and sequential data is significantly improved by deep learning methodologies, including Convolutional Neural Networks (CNNs), Recurrent Neural Networks (RNNs), Long Short-Term Memory (LSTM) networks, and Autoencoders. This results in more precise and contextually pertinent music recommendations. Moreover, hybrid models that amalgamate deep learning with conventional techniques augment recommendation precision by synthesising user interaction data with acoustic characteristics. This paper examines essential performance metrics employed in the assessment of music recommendation systems, including precision, recall, F1-score, and Mean Reciprocal Rank (MRR). It also tackles issues including computational complexity, bias, and ethical dilemmas pertaining to data privacy.

Keywords: Music recommendation, deep learning, Convolutional Neural Networks, Recurrent Neural Networks, hybrid models.

1. Introduction

In recent years, music recommendation systems have revolutionised the method in which people discover and interact with music, particularly on platforms such as Spotify, Apple Music, and YouTube. These services employ recommendation technologies to create tailored playlists and improve user engagement. Conventional techniques, such as collaborative filtering and content-based filtering, evaluate user behaviour or item characteristics for recommendations. Nonetheless, these methods encounter obstacles such as the cold-start problem and the challenge of accurately capturing the intricate correlations between user preferences and musical attributes, necessitating a transition to deep learning-based methodologies.

Deep learning, a branch of machine learning, has facilitated progress in areas such as computer vision and natural language processing, and is currently being utilised in recommendation systems. Its capacity to autonomously extract hierarchical features from unprocessed data renders it optimal for managing the diverse and intricate data associated with music suggestions. Music data is multi-modal, encompassing

DOI: 10.54254/2755-2721/109/20241353

audio signals, metadata, user interactions, and text-based information such as lyrics or reviews. The capacity of deep learning to analyse various data enables it to provide more tailored recommendations.

The present status of research in this domain has witnessed diverse methodologies for implementing deep learning in music recommendation. Schedl [1] performed a survey that emphasises the advancement of deep learning models in recommendation systems, particularly their proficiency in managing high-dimensional data in music. A pivotal study by Singh et al. [2] presented a hybrid deep learning model that amalgamates user preferences with audio information to improve suggestion precision. Furthermore, autoencoders have been investigated for mitigating the cold-start problem, with Singh et al. illustrating its effectiveness in managing sparse data via compact latent representations. The research encompasses the creation of emotion-aware systems, with Joshi et al. [3] examining the utilisation of deep learning for real-time mood detection to enhance personalised suggestions.

The increasing interest in utilising deep learning for music recommendation systems arises from numerous significant advantages. Convolutional Neural Networks (CNNs) and Recurrent Neural Networks (RNNs) excel in processing sequential and high-dimensional data. CNNs are extensively employed to extract features from audio spectrograms, thereby expressing music songs in a machine-readable format. CNNs improve recommendation precision by recognising patterns in both frequency and temporal domains. RNNs and Long Short-Term Memory (LSTMs) are proficient at identifying temporal dependencies in user listening behaviour, enabling systems to recommend music based on a user's interaction history.

Autoencoders and Variational Autoencoders (VAEs) offer an additional benefit by mitigating the cold-start difficulty. Conventional techniques such as collaborative filtering face challenges in promoting novel devices or accommodating new users because of insufficient data. Deep learning models, especially VAEs, may effectively manage latent features derived from sparse data, hence facilitating precise suggestions in cold-start scenarios. Hybrid methodologies that integrate deep learning with conventional techniques have arisen as effective solutions, improving both recommendation precision and scalability.

This research paper presents a survey of the diverse deep learning techniques utilised in music recommendation systems. The emphasis is on how various methodologies tackle distinct issues in the domain, including CNNs, RNNs, and hybrid models, assessing their efficacy regarding recommendation accuracy, scalability, and user happiness. This analysis analyses recent research to identify patterns and prospective advancements in music recommendation systems.

2. Background and Evolution of Music Recommendation Systems

Music recommendation systems are essential to the success of online music platforms such as Spotify, Apple Music, and YouTube. Initially, these systems employed conventional techniques such as collaborative filtering and content-based filtering. Collaborative filtering is predicated on the premise that people with analogous preferences will appreciate identical goods, whereas content-based filtering suggests items based on the attributes of previously favoured content. While these methodologies shown efficacy in the initial phases of recommendation systems, they faced considerable constraints, such as scaling challenges, the cold-start dilemma, and complications in managing intricate data relationships [4,5].

The cold-start problem is especially common in music recommendation systems, because new users and new tunes do not provide enough data for conventional algorithms to produce precise recommendations. The substantial number of accessible contents presents scaling issues, as traditional methods fail to keep up with the dynamic and expanding datasets of millions of people and songs [4,6].

To mitigate these constraints, researchers employed hybrid recommendation systems that integrate collaborative filtering with content-based methods. Although these hybrid systems represented progress, they remained inadequate in encapsulating the intricacies of human preferences, particularly as music consumption has evolved to be increasingly multi-modal, encompassing audio characteristics, metadata, and user interactions [6-8]. As a result, deep learning has emerged as a viable method to address these difficulties.

DOI: 10.54254/2755-2721/109/20241353

3. Deep Learning Approaches in Music Recommendation

3.1. Convolutional Neural Networks

CNNs have gained prominence in the music recommendation space due to their ability to process raw audio data effectively. One of CNNs' key strengths is their capacity to extract meaningful patterns from structured inputs, such as Mel spectrograms, which are time-frequency representations of audio tracks. These spectrograms are vital for modeling complex relationships between various audio features such as rhythm, timbre, and harmony. By applying convolutional layers to Mel spectrograms, CNNs can detect localized patterns in the audio signal and create insightful predictions regarding the genre, mood, or style of music [6,9].

For instance, a study that utilised CNNs and Mel spectrograms to recommend music tracks revealed that deep networks outperformed conventional methods in terms of recommendation accuracy. The convolutional layers extract both low- and high-level features from the audio data, which are subsequently transmitted through fully connected layers to predict user preferences [6]. This method enhances the quality of recommendations and resolves the cold-start issue by capitalising on the inherent structure of audio data, rather than exclusively relying on user history.

Furthermore, CNNs have the capacity to generalise across various domains. CNNs are a versatile instrument for music recommendation tasks, including the identification of user preferences and the classification of music into genres, as the same architecture used in image classification tasks can be adapted for audio data by modifying the input features [10].

3.2. Recurrent Neural Networks and Long Short-Term Memory

Although CNNs excel in extracting spatial features from audio data, RNNs and LSTM networks are more proficient in capturing temporal dependencies in user behaviour. RNNs are proficient in modelling sequential data, rendering them suitable for examining a user's listening history over time. User preferences in music recommendation frequently exhibit temporal patterns, with specific genres or artists favoured at different times of day, or users inclined towards certain moods in their playlists [5].

RNNs can represent temporal dynamics by preserving a hidden state that progresses with each input in the sequence. Traditional RNNs, however, are constrained by short-term memory limits and the vanishing gradient problem, hindering their ability to capture long-term dependencies. LSTM networks, a subtype of RNNs, address this challenge by integrating memory cells that preserve information across extended sequences. LSTM-based models have been utilised in music recommendation tasks to effectively capture both short-term and long-term consumer preferences [6].

LSTMs can forecast the subsequent song a user may choose to listen to based on their listening history, considering aspects such as recently played tracks, genre transitions, and prior interaction patterns. This improves the system's capacity to suggest music that corresponds with the user's present listening context and emotional state [4].

3.3. Autoencoders and Variational Autoencoders

Autoencoders serve as a potent tool in deep learning-driven recommendation systems, especially concerning data sparsity and the cold-start dilemma. Autoencoders are neural networks that learn to compress and rebuild input data, successfully capturing the most important properties in a lower-dimensional latent space. Autoencoders can obtain latent representations of users and objects in music recommendation, enabling the system to produce recommendations with limited data [5].

VAEs enhance this notion by integrating a probabilistic element, allowing them to model uncertainty within the latent space. This characteristic makes VAEs particularly beneficial for handling the intrinsic diversity of musical data and user preferences. VAEs can improve recommendation quality in cold-start situations by developing novel user-item interactions and addressing gaps in sparse datasets through the learning of a probabilistic distribution across latent variables [2,4].

Autoencoders are utilised in collaborative filtering systems to obtain succinct representations of user-item interactions, exemplifying a practical application of autoencoders in music recommendation. The

DOI: 10.54254/2755-2721/109/20241353

system's power to generalise across a broad user base is augmented by these latent traits' ability to predict a user's enjoyment of a certain song, even if they have never before listened to it [3].

3.4. Hybrid Models

Hybrid recommendation systems enhance accuracy and mitigate specific constraints by integrating the advantages of deep learning models with conventional recommendation methodologies. For instance, a hybrid model could combine collaborative filtering with a CNN-based feature extractor to capitalise on both auditory features and user interaction data. This method guarantees that the system can suggest tracks that are indicative of the music's inherent qualities and the user's listening history [6].

A notable example is the hybrid deep learning-based music recommendation system, which improves the accuracy of recommendations by integrating matrix factorisation techniques with deep neural networks. This system employs CNNs to extract features from auditory data, which are subsequently combined with user-item interaction data in a matrix factorisation framework. The outcome is a more personalised recommendation that takes into account the user's prior behaviour and the content of the music [5].

4. Performance Metrics and Evaluation

The review of music recommendation systems is essential for gauging their efficacy in delivering precise and pertinent music recommendations. Diverse performance criteria are employed to assess the efficacy of these systems in predicting user preferences and the satisfaction level of the recommendations provided to users. Precision, recall, F1-score, and Mean Reciprocal Rank (MRR) are among the most commonly utilised measures in recommendation systems, each providing unique insights into the quality of recommendations [4].

Precision is a metric that quantifies the ratio of pertinent items to the overall recommendations provided by the system. In music suggestion, a high precision score signifies that the majority of tunes suggested to the user are genuinely appreciated or pertinent, offering a direct assessment of correctness. If a system recommends 10 songs and 8 are pertinent, the precision would be 0.8 [5]. Precision, however, does not consider the system's capacity to identify all pertinent songs, highlighting the need of memory.

Recall denotes the system's capacity to recognise all pertinent items among the complete collection of available tunes. A high recall score indicates that the system effectively identifies a broad spectrum of pertinent songs, including ones that the user may not have previously indicated an interest in [6]. If the system identifies 20 songs that align with the user's preferences and accurately recommends 15, the recall would be 0.75. Nonetheless, a model with strong recall may nonetheless recommend irrelevant content, necessitating a balance with precision.

The F1-score is frequently utilised to reconcile the trade-off between precision and recall. The F1-score is the harmonic mean of precision and recall, offering a unified metric that integrates both accuracy and completeness. An elevated F1-score signifies that the system has attained equilibrium between precision and recall, providing a more thorough assessment of its efficacy in music recommendation [4].

The Mean Reciprocal Rank (MRR) is another significant indicator that emphasises the ranking of pertinent recommendations offered by the algorithm. The MRR quantifies the speed at which the system delivers a pertinent item to the user, assigning greater importance to items that are positioned earlier in the suggestion list. This is especially significant in music recommendation systems, as consumers are more inclined to engage with songs positioned near the top of the list, rather than perusing an extensive array of options. MRR is determined by averaging the reciprocal ranks of pertinent items, so guaranteeing that suggestions are both precise and suitably rated [5].

In addition to these quantitative indicators, user satisfaction is a crucial factor in assessing the efficacy of a music recommendation system. Although precision and recall offer objective performance metrics, user happiness reflects the subjective experience of the listener. This can be evaluated by surveys, retention rates, and the duration of user engagement with the suggested material. Research indicates that deep learning systems generally yield higher user satisfaction than conventional techniques, as they provide more personalised and contextually pertinent recommendations [6].

DOI: 10.54254/2755-2721/109/20241353

Recent research indicates that hybrid models have surpassed traditional methods across various evaluation parameters, especially in music recommendation tasks. Systems that integrate content-based filtering with deep learning methodologies, including CNNs and LSTMs, have demonstrated superior precision, recall, and F1-scores compared to models that depend only on collaborative filtering [2]. The use of deep learning into recommendation systems has improved the user experience by delivering more personalised suggestions derived from user interaction data, acoustic characteristics, and contextual elements.

Furthermore, sophisticated A/B testing is frequently employed to assess the efficacy of music recommendation systems in practical environments. A/B testing is the comparison of two iterations of a recommendation algorithm—one incorporating a novel feature (e.g., a deep learning model) and the other employing a baseline approach (e.g., collaborative filtering)—to evaluate which version yields superior user engagement and satisfaction. The outcomes of these assessments offer significant insights for enhancing recommendation algorithms [4].

Finally, scalability is a critical factor to evaluate in extensive music recommendation systems. Given that music platforms manage millions of tracks and users, it is imperative for recommendation algorithms to expand effectively while preserving optimal performance. Deep learning models, despite their efficacy, can be resource-intensive, and the evaluation process includes optimising these models to manage extensive datasets without sacrificing the quality of suggestions [5].

5. Challenges and Future Directions

Despite considerable progress, numerous problems remain in the implementation of deep learning for music recommendation systems. One of the primary issues is the computational complexity and scalability of these models. Training deep learning models on extensive datasets, such as those available on services such as Spotify or Apple Music, necessitates considerable computer resources and time. These platforms support millions of songs and users, complicating the situation for smaller platforms that may lack the necessary power to support such models. Moreover, the need for extensive hyperparameter optimisation increases resource demands and may delay practical deployment [5]. Researchers are exploring techniques to improve these models, including model pruning, quantisation, and knowledge distillation, to reduce training times and model sizes while preserving performance [6].

A significant difficulty is bias in recommendation systems. Deep learning algorithms trained on historical data often perpetuate biases that favour prominent artists and mainstream content, thereby marginalising less popular or specialist genres. This results in recommendations that lack diversity and may adversely affect both users and young artists. Future research should concentrate on integrating fairness restrictions into deep learning models. Models might be developed to equilibrate suggestions between mainstream and obscure recordings, thereby ensuring a more equal distribution of musical content. Furthermore, including user feedback enables the system to adapt its recommendations dynamically over time, hence minimising bias [4].

Ethical issues with data privacy and transparency have garnered attention, as music platforms accumulate extensive personal data, encompassing listening history, demographics, and emotional state. Users frequently lack awareness regarding the utilisation of their data, hence prompting concerns about consent. Future research should concentrate on the development of explainable AI (XAI) models. These models can offer users explicit justifications for specific recommendations, so enhancing trust and transparency [2].

A notable advancement in music recommendation is the emergence of emotion-aware systems. These systems employ deep learning to assess real-time emotional data, customising music recommendations to align with a user's present state. Such systems can provide more personalised and contextually appropriate music by recognising emotions through facial expressions, vocal intonations, or biosensors [1,3]. While these technologies improve user experience, they also provoke privacy concerns related to sensitive emotional data, necessitating a balance between innovation and ethical precautions.

A burgeoning concept is multimodal learning, which integrates many data types—such as auditory characteristics, user interactions, lyrics, and social context—to enhance the accuracy of suggestions.

DOI: 10.54254/2755-2721/109/20241353

Multimodal systems provide a more thorough suggestion process by examining a wider range of characteristics that affect user preferences [4].

Ultimately, cross-domain recommendation systems signify a promising domain for future investigation. These systems suggest items from one domain (e.g., music) depending on the user's behaviour in another domain (e.g., movies or books). By integrating user preferences across several domains, these systems offer more diversified and compelling recommendations. Implementing cross-domain recommendations necessitates deep learning models adept at managing disparate data sources [2].

6. Conclusion

The paper reviewed and integrated the principal breakthroughs in utilising deep learning for music recommendation systems, illustrating its efficacy in overcoming the shortcomings of conventional recommendation techniques including collaborative filtering and content-based filtering. Deep learning models, specifically CNNs, RNNs, LSTM networks, and Autoencoders, facilitate the efficient processing of intricate, multi-modal data such as audio features, user interaction history, and text. These models provide a more profound comprehension of user preferences and improve the precision of music recommendations, especially in scenarios where conventional algorithms falter, such as the cold-start problem.

A primary result of this synthesis is the focus on hybrid models that integrate the advantages of deep learning with conventional recommendation methods. These systems are more adept at utilising content-based features in conjunction with collaborative filtering, resulting in enhanced scalability and superior suggestion quality. The amalgamation of user data and musical content via deep learning enhances personalisation, rendering recommendations more pertinent to individual preferences.

Nonetheless, despite the advantages, obstacles remain, especially with computational complexity, scalability, prejudice, and ethical issues. Deep learning models, however potent, necessitate substantial resources, rendering large-scale implementation problematic. Moreover, biases in the training data may result in inequitable suggestions, privileging mainstream tracks over obscure content. Ethical considerations, such as user data protection and transparency in recommendation algorithms, necessitate additional focus.

Future research should prioritise the development of more efficient, scalable models that minimise resource consumption, while also resolving bias and privacy concerns. Emotion-aware systems, XAI, and cross-domain recommendations represent interesting avenues for the personalisation and enhancement of user experiences. By harmonising technology progress with ethical principles, the forthcoming generation of music recommendation systems will probably provide more personalised, transparent, and equitable solutions, enhancing the music discovery experience for consumers globally.

References

[1] Schedl, M. (2019). Deep learning in music recommendation systems. Frontiers in Applied Mathematics and Statistics, 5, 457883.

[2] Singh, J., Sajid, M., Yadav, C. S., Singh, S. S., & Saini, M. (2022, April). A novel deep neural-based music recommendation method considering user and song data. In 2022 6th International Conference on Trends in Electronics and Informatics (ICOEI) (pp. 1-7). IEEE.

[3] Joshi, S., Jain, T., & Nair, N. (2021, July). Emotion based music recommendation system using LSTM-CNN architecture. In 2021 12th International Conference on Computing Communication and Networking Technologies (ICCCNT) (pp. 01-06). IEEE.

[4] Maheshwari, C. (2023). Music Recommendation on Spotify using Deep Learning. arXiv preprint arXiv:2312.10079.

[5] Sunitha, M., Adilakshmi, T., & Unissa, M. (2022). Hybrid deep learning-based music recommendation system. In Computer Networks, Big Data and IoT: Proceedings of ICCBI 2021 (pp. 517-530). Singapore: Springer Nature Singapore.

DOI: 10.54254/2755-2721/109/20241353

[6] Yin, T. (2023). Music Track Recommendation Using Deep-CNN and Mel Spectrograms. Mobile Networks and Applications, 1-8.
[7] Jacobson, K., Murali, V., Newett, E., Whitman, B., & Yon, R. (2016, September). Music personalization at Spotify. In Proceedings of the 10th ACM Conference on Recommender Systems (pp. 373-373).
[8] Oramas, S., Nieto, O., Sordo, M., & Serra, X. (2017, August). A deep multimodal approach for cold-start music recommendation. In Proceedings of the 2nd workshop on deep learning for recommender systems (pp. 32-37).
[9] Elbir, A., & Aydin, N. (2020). Music genre classification and music recommendation by using deep learning. Electronics Letters, 56(12), 627-629.
[10] Saraswat, M. (2024, February). Music Recommendation System Using Deep Learning and Machine Learning. In 2024 IEEE International Conference on Computing, Power and Communication Technologies (IC2PCT) (Vol. 5, pp. 1782-1787). IEEE.