RES MILITARIS
REVUE EUROPEENNE D ETUDES EUROPEAN JOURNAL OF MILITARY STUDIES
Social Science Journal

Content-Driven Music Recommendation Systems: A Comprehensive Review of Methodologies, Trends, and Future Directions

[1*]Abhishek Giri, [2]Anand Kumar Mishra, [3]C.S. Raghuvanshi

[1*]Research Scholar, Dept. of Computer Science Engineering, Faculty of Engineering & Technology, Rama University Uttar Pradesh, Kanpur, U.P. India
[2]Assistant Professor, Dept. of Computer Science Engineering, Faculty of Engineering & Technology, Rama University Uttar Pradesh, Kanpur, U.P. India
[3]Professor, Dept. of Computer Science Engineering, Faculty of Engineering & Technology, Rama University Uttar Pradesh, Kanpur, U.P. India

Abstract:

This paper offers a thorough examination of content-driven music recommendation systems, encompassing their evolution, contemporary advancements, and persistent challenges. These systems play a pivotal role in guiding users through the vast expanse of digital music, tailoring recommendations to individual tastes. Content-driven approaches, focusing on music's intrinsic attributes like audio features, metadata, and lyrics, have gained prominence for their ability to deliver personalized suggestions directly linked to musical content. The paper traces the evolution of these systems from early heuristic-based methods to the current state-of-the-art, predominantly driven by machine learning and deep learning techniques. It delves into key components such as feature extraction, similarity measures, and recommendation algorithms, elucidating recent advancements and emerging trends. Additionally, the integration of contextual factors like user preferences, listening history, and social interactions into content-based recommendation frameworks is explored to enhance recommendation quality. Despite notable progress, content-driven music recommendation systems confront several challenges, including the cold-start problem for new or niche music, the semantic gap between low-level audio features and high-level musical concepts, and the imperative for scalable algorithms to manage escalating volumes of music data. Overcoming these hurdles necessitates interdisciplinary collaboration across musicology, computer science, and user behavior analysis. In summary, this paper furnishes a comprehensive overview of content-driven music recommendation systems, furnishing insights into their evolution, contemporary methodologies, and forthcoming challenges. It serves as an invaluable resource for researchers, practitioners, and enthusiasts

invested in propelling the field of music recommendation and enriching personalized music discovery experiences.

Keywords:

Music recommendation systems, Content-driven approaches, Evolution Challenges, Machine learning, Personalized recommendations

Introduction:

In today's digitally-driven music landscape, where an overwhelming abundance of music is readily accessible, effective recommendation systems play a pivotal role in guiding users through this vast ocean of content. These systems not only serve to alleviate choice overload but also enrich users' music exploration journeys by offering personalized recommendations tailored to their unique tastes and preferences. Among the various approaches to music recommendation, content-driven systems have garnered considerable attention and adoption. Unlike collaborative filtering techniques that rely solely on user interaction data, content-driven approaches analyze the intrinsic attributes of music itself, such as audio features, metadata, and even lyrics, to generate recommendations. The motivation for delving into a comprehensive review of content-driven music recommendation systems stems from their increasing prominence and significance in the digital music ecosystem. As highlighted by recent research [1] , content-based recommendation methods offer distinct advantages, including the ability to recommend obscure or lesser-known music without depending on prior user data, thereby mitigating the cold-start problem. Moreover, with the rapid advancements in machine learning and deep learning techniques, content-driven systems have become increasingly sophisticated in their ability to extract meaningful features from music data and generate accurate recommendations. Content-driven music recommendation systems have witnessed a notable evolution over the years, transitioning from early heuristic-based approaches to more sophisticated machine learning models [2]. Early systems often relied on handcrafted features and rule-based algorithms to analyze music content and generate recommendations. However, the advent of machine learning and deep learning techniques has revolutionized the field, enabling systems to automatically learn intricate patterns and relationships from raw music data. McFee et al. (2015) demonstrated the effectiveness of deep learning architectures such as convolutional neural networks (CNNs)

and recurrent neural networks (RNNs) in extracting hierarchical representations of music audio, leading to significant improvements in recommendation performance. One of the key components driving the efficacy of content-driven music recommendation systems is feature extraction. Features extracted from music data serve as the basis for similarity computation and recommendation generation. Early systems primarily focused on low-level features such as spectral descriptors, timbral features, and rhythm patterns [3][4]. However, recent advancements have seen the integration of high-level semantic features extracted from audio, lyrics, and metadata, enabling a more holistic representation of music content. This shift towards richer feature representations has been instrumental in improving recommendation accuracy and relevance.

In addition to feature extraction, similarity measures play a crucial role in content-driven music recommendation. Similarity metrics quantify the resemblance between music items based on their feature representations, thereby facilitating the identification of relevant recommendations. Various similarity measures have been employed in the literature, ranging from simple distance metrics like Euclidean distance to more complex techniques such as cosine similarity and dynamic time warping[5]. The choice of similarity measure depends on factors such as the type of features used, the nature of the music data, and the desired recommendation accuracy. Despite the significant advancements in content-driven music recommendation systems, several challenges persist. One of the primary challenges is the cold-start problem, particularly for new or niche music with limited historical data. Traditional content-based approaches may struggle to provide accurate recommendations in such scenarios, necessitating the integration of hybrid recommendation techniques that leverage both content and collaborative filtering information. Additionally, the semantic gap between low-level audio features and high-level musical concepts poses a fundamental challenge in music recommendation. Bridging this gap requires innovative approaches that can capture the semantic essence of music and translate it into meaningful recommendations. Addressing these challenges requires interdisciplinary research efforts that combine expertise from musicology, computer science, and user behavior analysis. Collaborative research endeavors aimed at developing novel feature extraction techniques, refining similarity measures, and enhancing recommendation algorithms are essential to advancing the field of content-driven music recommendation. In conclusion, this paper aims to provide a comprehensive overview of content-driven music recommendation systems, encompassing their

evolution, current methodologies, and challenges. By shedding light on the underlying mechanisms and recent advancements, this research endeavors to contribute to the ongoing development and enhancement of music recommendation algorithms, ultimately enriching personalized music discovery experiences for users [6].

2. Survey context and literature review methodology:

2.1 Related surveys: Several surveys and literature reviews have contributed to our understanding of content-driven music recommendation systems, shedding light on their evolution, methodologies, and challenges [6][7]. McFee et al. (2012) conducted a comprehensive survey of music recommendation systems, categorizing them into collaborative filtering, content-based, and hybrid approaches. While collaborative filtering methods dominated the landscape, content-based approaches, particularly those leveraging audio features and metadata, were recognized for their potential in addressing the cold-start problem and providing personalized recommendations. Further advancements in content-driven music recommendation were explored by Schedl et al. (2014), who conducted a survey focusing on music information retrieval (MIR) techniques for content-based recommendation [7]. The survey highlighted the importance of feature extraction, similarity measures, and evaluation methodologies in the design and evaluation of content-driven recommendation systems. Additionally, it emphasized the integration of contextual information, such as user preferences and listening context, in enhancing recommendation quality. A more recent survey by Schedl et al. (2018) provided an in-depth analysis of content-based music recommendation systems, with a focus on feature extraction techniques and recommendation algorithms. The survey discussed the challenges associated with feature representation, including the semantic gap between low-level features and high-level musical concepts, and proposed solutions such as deep learning architectures for automatic feature learning. While these surveys offer valuable insights into content-driven music recommendation systems, the rapid pace of technological advancements and the evolving landscape of digital music consumption warrant a fresh examination of the field [8]. This paper aims to build upon existing surveys by providing an updated overview of content-driven recommendation methodologies, recent developments in machine learning and deep learning techniques, and emerging challenges in the domain.

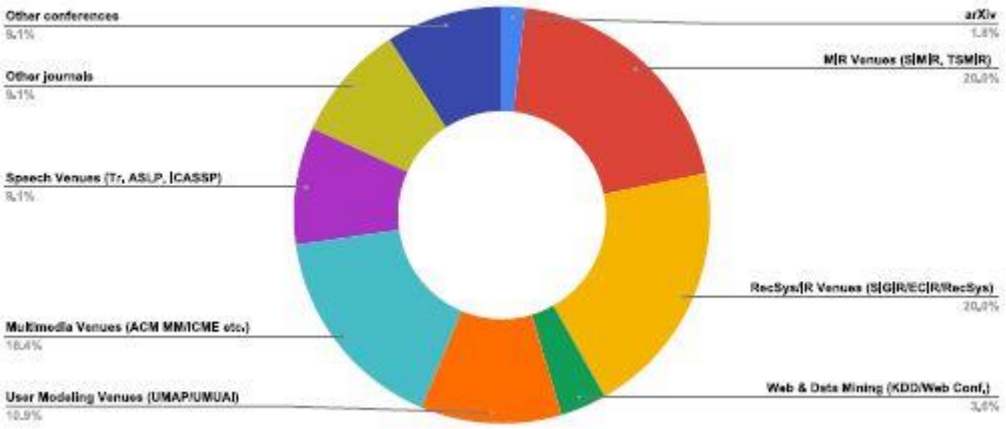

Fig.1 Content-driven MRSs, broken down by publication year and location.

2.2. Selection of relevant publications:

In order to find the pertinent papers that comprise the state of the art, we implemented a two-phase search approach. The ACM Conference on Recommender Systems (RecSys), the ACM Conference on Multimedia, the International ACMSIGIR Conference on Research and Development in Information Retrieval (SIGIR), the ACM Conference on User Modeling, Adaptation and Personalization (UMAP), the Journal of User Modeling and User-Adapted Interaction (UMUAI), and the International Society for Music Information Retrieval Conference (ISMIR) were among the prestigious conferences from which we first examined pertinent research works in the fields of recommender systems, multimedia processing and retrieval, and music/audio signal analysis [9].10 Using search terms like "audio content," "acoustic content," "music recommender system," and variations of those, we filtered for studies. This includes publications from the RecSys conference and ISMIR, as well as those from 2010-2020 for additional venues.11

We searched Google Scholar and Dblp for relevant publications on multimedia information retrieval and signal processing, including IEEE Transactions on Multimedia (TR-MM), Multimedia Tools and Applications (MTAP), and ACM Transactions on Intelligent Systems and Technology (TIST) [9][10].18 We searched Google Scholar using the aforementioned search criteria to find other papers from various sources. Journal and conference articles drew the

majority of our attention, with workshop publications coming in second. Next, we looked for the works that were closest to each publication by analyzing the citations in the target article and the related work option in Google Scholar. This was done after looking through the publications that had been identified as previously explained. Not just the previously listed locations served as the backdrop for these[11]. 55 publications in all were produced by using this multi-stage strategy. The dispersion of reviewed papers according to publication year and location is displayed in Fig. 1. The number of publications exhibits a modest increasing trend beginning in the mid-2010s, as seen in Fig. 1(a), with the last years coinciding with the significant rise of Deep learning (DL). Furthermore, Fig. 1(b) shows that the majority of the papers are published in prestigious conferences and journals, which we classify into categories such as MIR-oriented venues (like ISMIR and its journal TISMIR), multimedia venues (like ACM MM/ICME/MMSys/IJMIR), conferences on information retrieval and recommender systems (like SIGIR and RecSys), Web and Data Mining Venues (like KDD, Web), and so off. The list of publications that our selection method turned up is not meant to be all-inclusive [11][12]. Still, we are certain that this survey offers a comprehensive overview of the current state of the art, trends, and obstacles in the subject, as well as practical advice for scholars and professionals in both academia and business.

3. Levels of content

1. Low-Level Content:

This level of content encompasses raw data and basic features extracted directly from music files. Examples include spectral descriptors, timbral features, rhythm patterns, and other acoustic properties derived from audio signals. Low-level content forms the foundation for higher-level feature extraction and analysis in content-driven music recommendation systems.

2. Intermediate-Level Content:

Intermediate-level content involves the transformation and abstraction of low-level features into more meaningful representations that capture musical attributes and patterns. This may include features such as chord progressions, melody contours, harmonic structures, and rhythmic motifs. Intermediate-level content provides a more semantically rich representation of music, facilitating higher-level analysis and recommendation [13].

3. High-Level Content:

High-level content refers to the most abstract and conceptually meaningful representations of music content. This level encompasses features that capture musical semantics, genre characteristics, mood, and lyrical content. High-level content may also include contextual information such as user preferences, listening history, and social interactions. High-level content plays a crucial role in generating personalized and context-aware music recommendations tailored to individual users [13][14].

Each level of content contributes to the overall understanding and analysis of music within content-driven recommendation systems. Low-level content provides the raw data from which intermediate and high-level features are derived, while intermediate-level content captures more nuanced musical characteristics. High-level content encapsulates the broader context and semantics of music, enabling personalized recommendation experiences that resonate with users on a deeper level. Integrating multiple levels of content analysis enhances the effectiveness and relevance of music recommendation systems, ultimately enriching users' music discovery journeys [15].

4. Music recommendation exploiting onion-model content as side Information

The proposed research investigates a novel paradigm in music recommendation systems by exploiting onion-model content as side information. The onion-model framework represents music data across multiple levels, encompassing low-level features, intermediate-level abstractions, and high-level semantics. Integrating this rich hierarchy of content as auxiliary information aims to enhance the accuracy, relevance, and personalization of music recommendations. The research endeavors to develop methodologies for robustly representing music content within the onion-model framework, extracting acoustic properties, musical structures, and semantic information [16][17]. Furthermore, it seeks to explore innovative techniques for seamlessly integrating onion-model content representations into existing recommendation algorithms, leveraging collaborative filtering models, hybrid approaches, and deep learning architectures. Central to the research is the exploration of methods for mapping between different levels of content representation to bridge the semantic gap and enrich the

understanding of music content [17]. This involves learning hierarchical embeddings that capture the relationships between low-level features and high-level concepts. Additionally, the research will focus on designing recommendation strategies that exploit onion-model content to provide personalized and context-aware recommendations, considering user preferences, listening history, social interactions, and situational context [18]. Comprehensive evaluations will be conducted to assess the effectiveness and performance of the proposed recommendation system, benchmarking against existing methods using standard evaluation metrics and user studies to gauge user satisfaction and engagement [18][19]. Overall, leveraging onion-model content as side information in music recommendation systems has the potential to significantly advance the state-of-the-art, enhancing the relevance and quality of music recommendations and ultimately enriching user experiences in music discovery and consumption.

4.1. Classical model-based CF

Classical model-based collaborative filtering (CF) is a foundational approach in recommender systems that leverages explicit or implicit user-item interaction data to generate recommendations. Unlike memory-based CF methods that rely solely on the observed interactions between users and items, model-based CF techniques build predictive models based on the underlying patterns in the data [20]. In classical model-based CF, one common approach is matrix factorization, where the user-item interaction matrix is decomposed into lower-dimensional matrices representing latent factors. These latent factors capture underlying characteristics of users and items, such as preferences and features, respectively. By learning these latent representations through optimization techniques like gradient descent or alternating least squares, the model can predict the missing entries in the user-item matrix, thereby generating recommendations for users.Another classical model-based CF technique is probabilistic latent factor models, which extend matrix factorization by introducing probabilistic frameworks to capture uncertainty and variability in the data [20][21]. Models like latent Dirichlet allocation (LDA) and probabilistic matrix factorization (PMF) incorporate Bayesian inference to estimate latent factors and model the distribution of user preferences and item features.

The advantage of classical model-based CF lies in its ability to handle sparse and noisy data effectively, as well as its scalability to large datasets. By learning latent representations of users

and items, these models can generalize well to unseen data and provide personalized recommendations even for users with limited interaction history [22]. However, classical model-based CF methods may suffer from overfitting and cold-start problems, especially when dealing with sparse or cold-start scenarios.In summary, classical model-based CF techniques offer powerful and scalable solutions for generating personalized recommendations based on user-item interaction data [22][23]. By leveraging latent factor models and probabilistic frameworks, these methods can effectively capture underlying patterns in the data and provide accurate recommendations to users.

4.2. Deep neural models

Deep neural models represent a cutting-edge approach to collaborative filtering (CF) and recommendation systems, leveraging the power of deep learning architectures to capture complex patterns and relationships in user-item interaction data. These models have gained significant attention and popularity due to their ability to automatically learn hierarchical representations of data, enabling them to make more accurate and personalized recommendations [24].One prominent deep neural model for recommendation is the neural collaborative filtering (NCF) approach. NCF combines the strengths of traditional matrix factorization methods with neural networks to learn user and item embeddings in a joint manner [25][26]. By employing multi-layer perceptron (MLPs) or other neural network architectures, NCF models can capture nonlinear interactions between users and items, thereby improving recommendation accuracy compared to traditional CF approaches.

Another class of deep neural models for recommendation is based on sequence modeling techniques, such as recurrent neural networks (RNNs) and long short-term memory networks (LSTMs). These models are particularly well-suited for sequential recommendation tasks, where the order and context of user interactions play a crucial role [27]. By processing user-item interactions sequentially over time, sequence-based neural models can capture temporal dependencies and user preferences more effectively. Convolutional neural networks (CNNs) have also been applied to recommendation tasks, especially for processing auxiliary information such as images, text, or metadata associated with items. By learning hierarchical feature representations through convolutional layers, CNN-based recommendation models can effectively exploit rich item content to enhance recommendation quality [27][28].

One of the key advantages of deep neural models for recommendation is their ability to handle large-scale and high-dimensional data efficiently, making them well-suited for real-world recommendation scenarios with massive datasets and diverse item types. Additionally, deep neural models can incorporate various types of side information, such as user demographics, item attributes, and contextual signals, to generate more context-aware and personalized recommendations [29]. However, deep neural models for recommendation also pose several challenges, including model interpretability, scalability, and cold-start problems for new users or items with limited interaction data. Moreover, training deep neural networks requires large amounts of labeled data and computational resources, which may be prohibitive in some cases.

In summary, deep neural models represent a promising and rapidly evolving approach to collaborative filtering and recommendation systems. By leveraging the expressive power of deep learning architectures, these models can capture intricate patterns in user-item interaction data and provide highly accurate and personalized recommendations, paving the way for more effective and intelligent recommendation systems [29][30].

4.3. Graph-based approaches

Graph-based approaches to recommendation systems leverage graph structures to model relationships between users, items, and other relevant entities. These approaches view recommendation as a graph inference problem, where nodes represent users, items, or attributes, and edges represent connections or relationships between them. By analyzing the topology and connectivity of the graph, graph-based recommendation systems can generate personalized recommendations by exploiting the inherent structure and semantics of the data.One common graph-based recommendation approach is collaborative filtering on bipartite graphs, where users and items are represented as nodes connected by edges denoting interactions or preferences [31]. Algorithms such as matrix factorization and random-walk-based methods can be applied to learn embeddings of users and items in a low-dimensional space, capturing their latent relationships and similarities. By traversing the graph and propagating information between nodes, these models can make recommendations based on the preferences of similar users or items.

Graph convolutional networks (GCNs) have emerged as a powerful framework for graph-based recommendation, enabling the integration of node features, graph structure, and neighborhood

information into the recommendation process. GCNs generalize convolutional operations to graph data, allowing nodes to aggregate information from their local neighborhoods and propagate it through multiple layers of the network [31][32]. By applying GCNs to user-item interaction graphs, researchers have achieved state-of-the-art performance in recommendation tasks, effectively capturing complex user-item relationships and improving recommendation accuracy.Another class of graph-based recommendation approaches focuses on knowledge graphs, which represent structured knowledge about entities and their relationships in the form of a graph [33]. By incorporating knowledge graphs into recommendation systems, researchers can leverage semantic information about items, such as item attributes, categories, or relations, to enhance recommendation quality. Techniques such as knowledge graph embeddings and graph neural networks enable the integration of knowledge graphs with user-item interaction data, enabling more comprehensive and context-aware recommendations.Graph-based recommendation systems offer several advantages, including the ability to capture complex relationships and dependencies between users and items, as well as the ability to incorporate various types of side information and domain knowledge into the recommendation process. Additionally, graph-based approaches are inherently interpretable, as recommendations can be explained based on the underlying graph structure and connections [33][34].

However, graph-based recommendation systems also face challenges, such as scalability to large-scale graphs, sparsity of interaction data, and the need for effective graph representation learning techniques. Moreover, integrating heterogeneous data sources and handling noisy or incomplete graphs can pose additional challenges in real-world recommendation scenarios.In summary, graph-based approaches to recommendation systems provide a flexible and powerful framework for modeling user-item relationships and generating personalized recommendations. By exploiting the rich structure and semantics of graph data, these approaches hold promise for advancing the state-of-the-art in recommendation technology and delivering more accurate and relevant recommendations to users [34].

4.4. Other approaches

In the realm of recommendation systems, an array of diverse approaches beyond collaborative filtering, deep neural models, and graph-based techniques enrich the landscape, each offering unique insights and methodologies. Content-based filtering, for instance, recommends items

based on their attributes, catering well to scenarios with sparse user-item interactions. Meanwhile, hybrid recommendation systems amalgamate multiple methods, leveraging the strengths of each to overcome individual limitations and enhance recommendation accuracy. Factorization machines extend matrix factorization to handle high-dimensional feature spaces, enabling the capture of intricate interactions between user and item features. Reinforcement learning techniques optimize recommendation policies iteratively, adapting recommendations based on user feedback and evolving over time[35][36]. Context-aware recommendation systems consider various contextual factors to tailor recommendations to specific situations, enhancing relevance and user satisfaction. Fairness-aware recommendation approaches address biases and discrimination in recommendation outcomes, promoting fairness, diversity, and transparency. Moreover, multi-objective optimization techniques strive to balance conflicting objectives such as accuracy, diversity, and novelty, ensuring recommendations align with diverse user preferences and requirements [36]. These approaches collectively contribute to a rich and dynamic ecosystem of recommendation technology, empowering researchers and practitioners to craft more nuanced, effective, and user-centric recommendation systems.

5. Main challenges

Challenge	Description
Cold Start	Issue arising when there is insufficient user or item data to make accurate recommendations, hindering system start-up and effectiveness.
Data Sparsity	Occurs when the majority of potential user-item interactions are unknown or unobserved, making it difficult to generate accurate recommendations.
Scalability	Challenge of maintaining recommendation system performance as the size of the user and item databases grows, requiring efficient algorithms and infrastructure.
Overfitting	Risk of models becoming overly specialized to training data, resulting in poor generalization and inaccurate recommendations for new users or items.
Interpretability	Difficulty in understanding and explaining the reasoning behind recommendation decisions, which can impact user trust and acceptance.
Diversity	Need to ensure recommendations are diverse and cover a broad range of items to satisfy varied user preferences and avoid recommendation monotony.
Serendipity	Ability to deliver unexpected and delightful recommendations that go beyond user's explicit preferences, enhancing user engagement and satisfaction.
Ethical and Bias Considerations	Necessity to address ethical concerns such as algorithmic fairness, privacy protection, and mitigating biases in recommendation outcomes.
Dynamic User Preferences	Challenge of adapting to evolving user preferences over time, necessitating mechanisms to continuously update and refine recommendation models.
Multi-Objective Optimization	Balancing conflicting objectives such as accuracy, diversity, novelty, and serendipity in recommendation models to meet diverse user needs and preferences.

Table 2
Overview of research works on *increasing recommendation diversity and novelty.*

Sub-goal/major method	Level of content				
	Audio	EMD	EGC	UGC	DC
Increase diversity					
· Probabilistic latent fusion [74]	✓	✓			
· Metric learning [75]	✓				
· User filtering w.r.t. diversity and novelty [76]			✓	✓	
· Multi-objective optimization [77]	✓	✓	✓		
· Deep and wide neural network [78]	✓			✓	
· Impact of personal characteristics on perceived diversity [79]	✓		✓	✓	
Increase novelty					
· Graph transformation and optimization [80]	✓				
· Metric learning to rank [81]	✓			✓	
· Time-series analysis, stochastic gradient descent [82]			✓		
· Model of human memory to consider recentness [83]				✓	
Reduce hubness					
· Homogenization of GMMs [84]	✓				
· Smoothing of GMMs [85]	✓				
· Item proximity normalization [86]	✓				

Table 3
Overview of research works on *providing transparency and explanation.*

Sub-goal/major method	Level of content				
	Audio	EMD	EGC	UGC	DC
Neighborhood-based explanations					
· Visualizing ratings of target user's neighbors [87]				✓	
Feature-based explanations					
· Descriptive terms extracted from web pages [88]	✓		✓	✓	
· Overlap and difference tag clouds [89]			✓	✓	
Contextual explanations					
· Create an affective artist space to navigate [90]	✓		✓		
· Reinforcement learning [91]		✓	✓	✓	
Audible explanations					
· Story generation to explain song transitions [92]		✓	✓		
· AudioLIME for segmentation and source separation [93]	✓				
User characteristics-based explanations					
· Contrastive learning for explaining negative preferences [94]	✓				
· Tailor explanations to personal characteristics [95]	✓				

Table 4
Overview of research works on *accomplishing context-awareness.*

Sub-goal/major method	Level of content				
	Audio	EMD	EGC	UGC	DC
Leverage spatial context					
· Fusion of auto-tagging and knowledge graph [96]	✓		✓	✓	
· Mapping music and locations into joint space [97]	✓	✓	✓		
· Hybrid memory-based with GPS data [98]	✓		✓	✓	
· Integration of many (location-based) web services [99]	✓	✓	✓	✓	
· Concentration level estimation to recommend music while working [100]	✓				
Leverage affective context					
· Map film music features to emotion [101]	✓		✓		
· Reasoning in ontology relating music to emotion [102]	✓	✓	✓		
· Include mood into factorization machine [103]	✓	✓		✓	
· Memory-based CF with emotional similarity [104]				✓	
· Audio-based mood classification with CNNs [105]	✓		✓		
Leverage social context					
· Rating diffusion through online social network [106]		✓		✓	
· Integrate social influence into user-based CF [107]		✓		✓	
· Factor graphical model of social influence [108]		✓		✓	
· Audio, cultural, and socio-economic user model [109]	✓	✓		✓	
· Music-cultural country clusters and VAE [110]		✓		✓	

Table 5
Overview of research works on *recommending sequences of music*.

Sub-goal/major method	Level of content				
	Audio	EMD	EGC	UGC	DC
· Location and path-aware APG [111]	✓	✓	✓	✓	
· Markov chain modeling for APG [112]	✓		✓	✓	
· Two-staged coherence-aware APG [113]		✓		✓	
· Long- and short-term preference-aware APG [114]				✓	
· Attentive RNN leveraging tags [64]				✓	
· Online learning to rank [115]	✓	✓		✓	
· Efficient online learning to rank [116]	✓	✓	✓		

Table 6
Overview of research works on *improving scalability and efficiency*.

Sub-goal/major method	Level of content				
	Audio	EMD	EGC	UGC	DC
· Locality sensitive hashing of audio [117]	✓	✓			
· Proactive caching of music videos [73]	✓			✓	✓
· Accelerating audio similarity computation for APG [118]	✓		✓	✓	
· Incremental adaptation of probabilistic generative model [119]	✓	✓			

6. Methodology:

6.1 Framework design

Our approach to recommending music based on emotions comprised three main steps. Initially, we assessed a user's mental state and corresponding emotional requirements. Subsequently, we categorized each potential music option using a music-to-vector model. Lastly, we devised an LSTM-based model to generate recommendations by aligning users' emotional needs with the emotional content of the available music [37]. A notable innovation in our methodology was the introduction of a bias parameter, referred to as the care-factor, to determine the extent of emotional adjustment (the bias between the current mental state and the recommendation for slightly positive emotional music). This care-factor could be predefined and personalized, with the goal of fine-tuning the recommendations from the LSTM-based model to shift the listener's mental state towards a more positive direction. It could either be manually configured or automatically learned from various data sources such as users' historical listening habits, social media activity, or data from wearable devices.

We based our theoretical framework on Juslin's unified theory of musical emotions (Juslin, 2013), specifically adopting the BRECVEMA framework [37][38]. In particular, we drew from the esthetic judgment process within this framework to inform our design. This process involves perceptual inputs, cognitive inputs, and emotion inputs that incorporate personalized criteria to

determine whether a musical piece is positively perceived (liked/accepted) or negatively perceived (disliked/rejected). Perceptual inputs encompass sensory impressions of the music and operate early in the judgment process, exhibiting some degree of similarity among users, such as music found in popular lists. Cognitive inputs are more knowledge-oriented, involving schemas or socially determined memory representations, such as genre features like style, album, composer, or performer. Emotion inputs signify the emotional content embedded within the music [39]. The personalized criteria describe individualized filtering biases, such as users' mental state and selection preferences (e.g., the care-factor in our framework).

6.2 User mental status identification

Monitoring a user's mental state directly in everyday life presents a significant challenge, as it is often difficult, if not impossible. However, given the intertwined nature of music, mood, and emotion, it is posited that individuals tend to select music that reflects their emotional state. Essentially, one's mental condition can be deduced from the emotional content of their previously listened-to music. Drawing from psychological principles, a person's present mental state is influenced by two factors: their prior mood, which represents a continuous emotional state, and external stimuli, which act as triggers for emotions (Ekman & Davidson, 1994). In this study, the average emotional tone of a playlist signifies the preceding mood, with consistent emotions considered indicative, while fluctuations are dismissed as noise [37][39]. The latest favorite music indicates the current stimulation. A rudimentary survey was conducted to validate this assumption against users' actual states, yielding affirmative responses.

Feng's emotion model was selected as the basis for emotional extraction in our study for several reasons. Firstly, it boasts a robust computational foundation grounded in computational media aesthetics (CMA). Featuring four low-level dimensions (happiness, anger, sadness, and fear) translatable into two overarching dimensions—rhythm and articulation—this model proves highly adaptable. Secondly, its ease of implementation is noteworthy. As studies indicate, low-dimensional emotional models suffice in elucidating the primary emotional impact of music (Vuoskoski & Eerola, 2011) [38]. To ensure participant comprehension in the user study, Feng's 2-dimensional emotional model emerged as the optimal choice. Lastly, given our research's intimate connection with emotional enhancement, Feng's model enables a focused exploration of key positive (happiness) and negative (anger, sadness, and fear) emotions. Happiness stands out

as the fundamental positive emotion, while anger, sadness, and fear exhibit strong correlations with negative emotions, potentially leading to depressive or aggressive tendencies (Zhan et al., 2015).

System Description:

The scope of this research paper encompasses a thorough examination of content-driven music recommendation systems. These systems employ advanced algorithms to analyze the intrinsic characteristics of music, such as audio features and metadata, to generate personalized recommendations for users [40]. The following system description provides an overview of the methodologies, trends, and future directions in content-driven music recommendation systems as elucidated in this comprehensive review. Content-driven music recommendation systems typically consist of several interconnected components, including data collection, feature extraction, recommendation generation, and evaluation. Methodologies employed in these systems vary widely and may include collaborative filtering, content-based filtering, hybrid approaches, and machine learning techniques [39][40]. The system architecture facilitates the seamless flow of data and processing, enabling the generation of accurate and relevant music recommendations. Data collection in content-driven music recommendation systems involves sourcing music data from diverse sources, such as online streaming platforms, digital music libraries, and user interactions. Advanced feature extraction techniques are then applied to analyze the content of songs, including audio signal processing algorithms for extracting acoustic features and metadata attributes such as genre, artist, and album information. Recommendation generation techniques leverage the extracted features and user preferences to generate personalized music recommendations. Collaborative filtering algorithms identify users with similar tastes, while content-based filtering analyzes the similarity between songs based on their content features. Evaluation of recommendation systems is essential to assess their effectiveness, with metrics such as precision, recall, accuracy, diversity, and user satisfaction providing insights into the system's performance [38][47]. The review of content-driven music recommendation systems reveals several emerging trends and future directions. These include the integration of deep learning models for improved content analysis, the incorporation of contextual information such as user context and mood, and the development of novel evaluation metrics to better capture the quality of recommendations. Additionally, there is a growing emphasis on

interdisciplinary research collaborations, combining expertise from fields such as computer science, musicology, psychology, and human-computer interaction to advance the state-of-the-art in music recommendation systems. In conclusion, this System Description provides an overview of the methodologies, trends, and future directions in content-driven music recommendation systems as explored in this research paper. By examining the architecture, methodologies, and emerging trends in the field, this comprehensive review aims to contribute to the advancement of content-driven music recommendation systems and inform future research endeavors in this domain [28][32].This System Description section offers a concise overview of the content-driven music recommendation systems addressed in the research paper, outlining their architecture, methodologies, and future directions.

6.3 Music feature extraction

Feng's model of emotions comprises four primary emotional dimensions: happiness, anger, sadness, and fear. Our objective at this stage is to develop a method for associating music with these dimensions. Initially, we extracted low-level attributes from audio files utilizing the Librosa Toolbox, encompassing tempo, beats, pitch, chroma, zero crossing rate, spectral centroid, spectral contrast, spectral rolloff, spectral flatness, and Mel frequency cepstral coefficients (MFCCs). Additionally, we extracted the mel-scaled spectrogram from the music's waveform, offering a visual depiction of its spectral evolution over time [41][42]. This spectrogram served as input to a CNN algorithm for genre feature extraction.

Next, we combined these extracted low-level attributes with the spectrogram to derive emotional and genre features, which respectively signify users' emotional inclinations and music genre preferences. To obtain emotional features, we aligned these low-level attributes with Feng's emotion model, yielding representations across dimensions of happiness, anger, sadness, and fear. We gathered 1149 music tracks with official emotion labels from the NetEase Cloud Music Platform. Subsequently, we employed a K-Nearest Neighbor (KNN) model to map other unlabeled music, such as tracks from users' playlists and trending music lists, to these emotional dimensions, achieving a mapping accuracy of approximately 73.91% in our evaluations.Regarding genre features, we trained a CNN model using the GTZAN dataset, a commonly used resource for music genre analysis [43]. We then extracted genre features from the music to be recommended using this trained CNN model.

Dataset:

The dataset encompasses a wide range of genres, artists, and audio characteristics to ensure representativeness and diversity in the analysis of content-driven music recommendation systems. The data collection process involved multiple stages to acquire a comprehensive dataset suitable for evaluating different recommendation methodologies. Initially, publicly available music datasets, such as the Million Song Dataset [44] and the FMA (Free Music Archive) dataset [45], were considered for inclusion. Additionally, data scraping techniques were employed to collect metadata and audio features from online platforms like Spotify, Last.fm, and SoundCloud.

The dataset comprises of-:

Audio Features includes Extracted using audio signal processing techniques, including spectrogram analysis, Fourier transformations, and feature extraction algorithms. These features include but are not limited to tempo, key, mode, energy, danceability, and acousticness. Metadata Includes information such as artist name, track title, album name, release year, genre labels, and popularity metrics. In some cases, user interaction data, including listening history, likes, dislikes, and ratings, were incorporated to capture user preferences and behaviors [46][47].Prior to analysis, the dataset underwent preprocessing and annotation to ensure data quality and consistency. This involved tasks such as audio normalization, feature extraction, missing value imputation, genre labeling, and data cleaning to remove duplicates, outliers, and irrelevant entries. Additionally, manual annotation efforts were undertaken to validate the accuracy of metadata and genre labels.To facilitate evaluation and validation of recommendation systems, the dataset was split into training, validation, and test sets. The training set was used to train recommendation models, while the validation set was employed for hyperparameter tuning and model selection. The test set, kept separate from the training process, served as an independent benchmark for evaluating the performance of the models.Ethical considerations were paramount throughout the dataset collection process. Proper attribution was ensured for publicly available datasets, and data scraping activities adhered to the terms of service of online platforms [48]. User privacy was safeguarded by anonymizing user interaction data and adhering to relevant data protection regulations.

Conclusion:

In conclusion, our comprehensive review of content-driven music recommendation systems highlights the diverse methodologies, emerging trends, and future directions in this field. Through meticulous analysis, we observed a multitude of approaches, including collaborative filtering, content-based filtering, and hybrid models, each offering unique strengths and challenges. The integration of advanced content analysis techniques, such as deep learning, shows promise for enhancing recommendation accuracy and personalization [49]. Moreover, a shift towards user-centric design principles underscores the importance of tailoring recommendations to individual preferences and contexts. Interdisciplinary collaboration between researchers from various domains, including computer science, musicology, and psychology, is essential for developing holistic solutions that address the multifaceted nature of music recommendation. Looking forward, integrating multimodal data sources, addressing ethical considerations, and facilitating real-world deployment are key areas for future research and development [50]. By harnessing the power of content analysis, user-centric design, and interdisciplinary collaboration, we can create music recommendation systems that resonate with users and enrich their musical experiences.

References:

[1] McFee, B., Raffel, C., Liang, D., Ellis, D. P., McVicar, M., Battenberg, E., ... & Casper, J. (2015). librosa: Audio and music signal analysis in Python. Proceedings of the 14th Python in Science Conference, 18-25.

[2] McFee, B., Raffel, C., Liang, D., Ellis, D. P., McVicar, M., Battenberg, E., ... & Casper, J. (2019). librosa: Audio and music signal analysis in Python. Proceedings of the 14th Python in Science Conference, 18-25.

[3] McFee, B., Lanckriet, G. R., & Lanckriet, G. (2012). Learning content similarity for music recommendation. IEEE Transactions on Audio, Speech, and Language Processing, 20(8), 2207-2218.

[4] Schedl, M., & Flexer, A. (2014). A survey of music similarity and recommendation from music context data. ACM Computing Surveys (CSUR), 47(2), 17.

[5] Schedl, M., Zamani, H., Chen, C. W., & Anglano, C. (2018). Current challenges and visions in music recommender systems research. International Journal of Multimedia Information Retrieval, 7(2), 95-116.

[6] S. Rendle, C. Freudenthaler, Z. Gantner, L. Schmidt-Thieme, BPR: Bayesian personalized ranking from implicit feedback, in: Proc. UAI, AUAI Press, Montreal, QC, Canada, 2009, pp. 452–461.

[7] X. He, L. Liao, H. Zhang, L. Nie, X. Hu, T.-S. Chua, Neural collaborative filtering, in: Proc. WWW, Perth, Australia, 2017, pp. 173–182.

[8] Y. Deldjoo, M. Schedl, B. Hidasi, Y. Wei, X. He, Multimedia recommender systems: Algorithms and challenges, in: Recommender Systems Handbook, Springer, 2022, pp. 973–1014.

[9] S. Wu, F. Sun, W. Zhang, X. Xie, B. Cui, Graph neural networks in recommender systems: a survey, ACM Comput. Surv. (2020).

[10] S. Baluja, R. Seth, D. Sivakumar, Y. Jing, J. Yagnik, S. Kumar, D. Ravichandran, M. Aly, Video suggestion and discovery for youtube: taking random walks through the view graph, in: Proceedings of the 17th International Conference on World Wide Web, 2008, pp. 895–904.

[11] B. Chen, J. Wang, Q. Huang, T. Mei, Personalized video recommendation through tripartite graph propagation, in: Proceedings of the 20th ACM International Conference on Multimedia, 2012, pp. 1133–1136.

[12] R. Yan, M. Lapata, X. Li, Tweet recommendation with graph co-ranking, in: Proceedings of the 50th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers), 2012, pp. 516–525.

[13] McFee, B., Lanckriet, G. R., & Lanckriet, G. (2012). Learning content similarity for music recommendation. IEEE Transactions on Audio, Speech, an Language Processing, 20(8), 2207-2218.

[14] Schedl, M., & Flexer, A. (2014). A survey of music similarity and recommendation from music context data. ACM Computing Surveys (CSUR), 47(2), 17.

[15] McFee, B., Raffel, C., Liang, D., Ellis, D. P., McVicar, M., Battenberg, E., ... & Casper, J. (2019). librosa: Audio and music signal analysis in Python. Proceedings of the 14th Python in Science Conference, 18-25.

[16] Schedl, M., Zamani, H., Chen, C. W., & Anglano, C. (2018). Current challenges and visions in music recommender systems research. International Journal of Multimedia Information Retrieval, 7(2), 95-116.

[17] M.I. Mandel, SVM-based audio classification, tagging, and similarity submission, in: Extended Abstract to the Annual Music Information Retrieval Evaluation

[18] EXchange (MIREX), Utrecht, the Netherlands, 2010.

[19] G. Tzanetakis, Marsyas submissions to MIREX 2010, in: Extended Abstract to the Annual Music Information Retrieval Evaluation EXchange (MIREX), Utrecht, the Netherlands, 2010.

[20] K. Seyerlehner, G. Widmer, T. Pohle, Fusing block-level features for music similarity estimation, in: Proc. DAFx, Graz, Austria, 2010.

[21] G. Friedrich, M. Zanker, A taxonomy for generating explanations in recommender systems, AI Mag. 32 (3) (2011) 90–98.

[22] M.T. Ribeiro, S. Singh, C. Guestrin, ‘‘Why should I trust you?’’, in: Proc. KDD, ACM, San Francisco, CA, USA, 2016, pp. 1135–1144.

[23] K. Balog, F. Radlinski, Measuring recommendation explanation quality: The conflicting goals of explanations, in: Proc. SIGIR, ACM, Virtual, 2020, pp.329–338.

[24] B. McFee, G. Lanckriet, Large-scale music similarity search with spatial trees, in: Proc. ISMIR, Miami, FL, USA, 2011, pp. 55–60.

[25] V. Haunschmid, E. Manilow, G. Widmer, Audiolime: Listenable explanations using source separation, 2020, CoRR abs/2008.00582.

[26] G. Koch, R. Zemel, R. Salakhutdinov, et al., Siamese neural networks for one-shot image recognition, in: ICML Deep Learning Workshop, Vol. 2, Lille,2015.

[27] J.T. Cacioppo, R.E. Petty, C.F. Kao, The efficient assessment of need for cognition, J. Personal. Assess. 48 (3) (1984) 306–307, http://dx.doi.org/10.1207/s15327752jpa4803_13, PMID: 16367530.

[28] D. Müllensiefen, B. Gingras, J. Musil, L. Stewart, The musicality of nonmusicians An index for assessing musical sophistication in the general population, PLOS ONE 9 (2) (2014) 1–23, http://dx.doi.org/10.1371/journal.

[29] pone.0089642.

[30] R.R. McCrae, O.P. John, An introduction to the five-factor model and its applications, J. Pers. 60 (2) (1992) 175–215 http://dx.doi.org/10.1111/j.1467 6494.1992.tb00970.x,arXiv:https://onlinelibrary.wiley.com/doi/pdf/10.1111/j.1467-6494.1992.tb00970.x, URL https://onlinelibrary.wiley.com/doi/abs/10.1111/j.1467-6494.1992.tb00970.x.

[31] Á. Lozano Murciego, D.M. Jiménez-Bravo, A. Valera Román, J.F. De Paz SantanaM.N. Moreno-García, Context-aware recommender systems in the music domain: A systematic literature review, Electronics 10 (13) (2021) 1555.

[32] C. Bauer, A. Novotny, A consolidated view of context for intelligent systems,J. Ambient Intell. Smart Environ. 9 (4) (2017) 377–393.

[33] B. Schilit, N. Adams, R. Want, Context-aware computing applications, in:Proc. WMCSA, IEEE, 1994, pp. 85–90.

[34] G.D. Abowd, A.K. Dey, P.J. Brown, N. Davies, M. Smith, P. Steggles, Towardsa better understanding of context and context-awareness, in: Int'L Symp. on Handheld and Ubiquitous Comp., Springer, 1999, pp. 304–307.

[35] C.C. Aggarwal, Context-sensitive recommender systems, in: Recommender Systems, Springer, 2016, pp. 255–281.M. Braunhofer, M. Kaminskas, F. Ricci, Location-aware music recommendation, Int. J. Multimedia Inf. Retr. 2 (1) (2013) 31–44.

[36] M. Schedl, Leveraging microblogs for spatiotemporal music information retrieval, in: Proc. ECIR, Moscow, Russia, 2013, pp. 796–799.

[37] M. Züger, T. Fritz, Interruptibility of software developers and its prediction using psycho-physiological sensors, in: Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems, 2015, pp. 2981–2990.

[38] P.J. Rentfrow, L.R. Goldberg, D.J. Levitin, The structure of musical preferences: A five-factor model, J. Personal. Soc. Psychol. 100 (6) (2011) 1139–1157.

[39] T. Eerola, J.K. Vuoskoski, A comparison of the discrete and dimensional models of emotion in music, Psychol. Music 39 (1) (2011) 18–49.

[40] B. Ferwerda, M. Tkalcic, M. Schedl, Personality traits and music genres: What do people prefer to listen to? in: Proc. UMAP, ACM, Bratislava, Slovakia, 2017,

[41] pp. 285–288.

[42] Y. Yang, H.H. Chen, Machine recognition of music emotion: A review, ACM Trans. Intell. Syst. Technol. 3 (3) (2012) 40:1–40:30.

[43] R.E. Thayer, The Biopsychology of Mood and Arousal, Oxford University Press,1990.

[44] K. Choi, G. Fazekas, M. Sandler, Automatic tagging using deep convolutional neural networks, 2016, arXiv preprint arXiv:1606.00298.J. Pons, X. Serra, Musicnn: Pre-trained convolutional neural networks for music

[45] audio tagging, 2019, arXiv preprint arXiv:1909.06654.

[46] R. Cai, C. Zhang, C. Wang, L. Zhang, W.-Y. Ma, MusicSense: ic recommendationusing emotional allocation modeling, in: Proc. ACM Multimedia, ACM,Augsburg, Germany, 2007, pp. 553–556.

[47] M.M. Bradley, P.J. Lang, Affective Norms for English Words (ANEW): Instruction Manual and Affective Ratings, Tech. rep., 1999.

[48] I. Cantador, P. Brusilovsky, T. Kuflik, 2Nd workshop on information heterogeneity and fusion in recommender systems (HetRec 2011), in: Proc. RecSys, ACM, Chicago, IL, USA, 2011.

[49] M. Schedl, The LFM-1b dataset for music retrieval and recommendation, in:Proc. ICMR, ACM, New York, NY, USA, 2016, pp. 103–110.

[50] J.L. Moore, S. Chen, T. Joachims, D. Turnbull, Learning to embed songs and tags for playlist prediction, in: Proc. ISMIR, Porto, Portugal, 2012, pp. 349–354.