Evolution and Emerging Trends in Musical Information Retrieval: A Comprehensive Review and Future Prospects

Yuxin Ding *

School Of Computing and Data Science, Xiamen University Malaysia, Xiamen, China

* Corresponding author: SWE2009496@xmu.edu.my

Abstract. The rapid development of the digital music industry has brought challenges for music lovers and researchers. In response to these challenges, the field of Music Information Retrieval (MIR) emerged in the mid-1960s to capture the complicated and multi-layered nature of music. Among the various approaches explored, machine learning methods have shown promise in overcoming the complexity of this interdisciplinary field. This paper is primarily centered on conducting an in-depth review of the current state of research regarding the utilization of machine learning within the field of Music Information Retrieval (MIR). Additionally, it aims to forecast the potential future directions within the MIR industry. Upon extensive literature review, it becomes evident that various machine learning techniques, including Neural Networks (NN), Support Vector Machines (SVM), and K-nearest neighbors (KNN), have found common applicability in this field. Furthermore, this review highlights Convolutional Neural Networks (CNN) and Support Vector Machines (SVM) as potential algorithms poised to shape the future landscape of MIR. The findings of this paper serve to elucidate the direction in which MIR is progressing, offering valuable guidance for forthcoming research and development endeavors. In doing so, it contributes to the continued progress and maturation of the field MIR.

Keywords: Musical Information Retrieval, Machine learning, CNN, SVM.

1. Introduction

The advent of the internet in the music industry has ushered in an era of information explosion in global music. The availability of music audio from around the world has given audiences unprecedented access to a vast amount of musical content. However, this abundance of musical data has created a significant challenge for music enthusiasts and researchers [1]. Music can be represented in various forms, including audio recordings, musical notations, and time-stamped events. Developing a single, universal method to encompass all these diverse musical expressions is a formidable task due to their distinct features. Furthermore, the simultaneous involvement of multiple instruments, including vocal elements, in music performance adds a layer of complexity to the analysis.

In response to these challenges, the field of Music Information Retrieval (MIR) emerged in the mid-1960s, attempting to encompass diverse musical expressions. However, the ambiguous nature of music and emotion poses a fundamental challenge. Music's definition is elusive, and individual listeners often have different emotional responses to the same piece. The complexity of music representation, the multifarious nature of musical instruments and expressions, and the nuanced interplay between music and emotion collectively contribute to the ongoing maturation of the MIR field.

Despite the considerable volume of research conducted in the field of MIR, it continues to grapple with these intricate and multifaceted challenges. The rapid evolution of digital music and music streaming services has given rise to the interdisciplinary field of MIR. MIR is primarily concerned with the research and development of algorithms and methodologies aimed at the automated extraction, analysis, and retrieval of information from music data. This field assumes a central role in various applications within the music industry, encompassing content-based music search, personalized playlist generation, music recommendation systems, and music discovery platforms.

In addressing the challenges confronted by MIR, machine learning methods have emerged as a compelling solution. Machine learning models can be effectively trained using extensive music

datasets, enabling them to provide predictions and recommendations based on vast music libraries. Additionally, machine learning excels in handling multimodal data, a critical feature in MIR. Furthermore, these models possess the capability to automatically derive relevant features, which proves invaluable in tasks such as music genre classification and music emotion recognition. Presently, MIR encompasses several subfields that leverage machine learning, including musical genre classification, musical instrument recognition, and prominently, music emotion recognition.

This study provides a thorough analysis of recent advancements in utilizing machine learning methodologies within MIR. It explores topics such as musical genre classification, musical instrument recognition, and music emotion recognition. Additionally, the study briefly analyzes emerging trends and challenges in the field, emphasizing the pivotal role of machine learning in the evolution of Music Information Retrieval.

2. Machine Learning Methods Applied in MIR

Within the broader domain of MIR, one well-established and essential task is music genre classification, which aims to categorize music into distinct genres based on its acoustic characteristics and content. To enhance computational efficiency and address issues related to premature convergence, innovative algorithms have been incorporated into the domain of music genre classification. For instance, the African Buffalo Optimization (ABO) method, inspired by the behavior of African buffalos, can remember previous solutions, encourage collaboration among optimization agents, and minimize the reliance on numerous parameters [2]. The ABO algorithm enhances optimization efficiency by leveraging these characteristics. The process flow of the ABO algorithm is shown in Fig.1 [2].

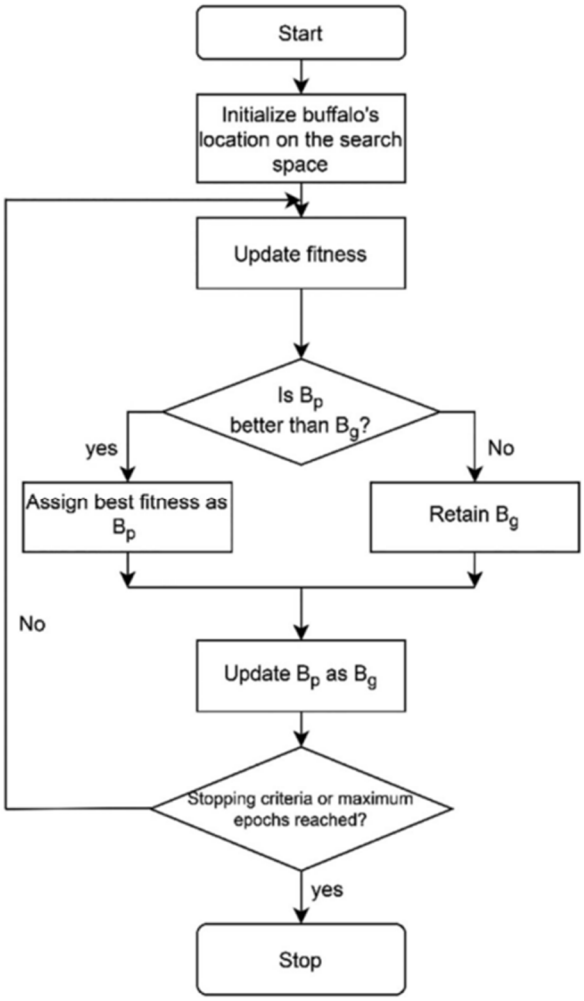

Figure 1. Process flow of African buffalo optimization [2].

In the domain of MIR, traditional algorithms, including Neural Networks (NN), Support Vector Machines (SVM), and K-nearest neighbors (KNN), have been commonly used. However, notable breakthroughs have emerged in recent years, particularly in July 2019, addressing a significant challenge in MIR, which is the limited availability of large-scale annotated data for supervised machine learning algorithms [3]. Additionally, Convolutional Neural Networks (CNN) have exhibited superior performance compared to SVM, and there is ongoing research into combining CNN and SVM to further enhance performance [4]. Singer Identification [3], a crucial component of MIR, can also be integrated into Musical Instrument Recognition.

Music Emotion Recognition [5], a rapidly evolving field within MIR, has witnessed substantial advancements in recent years, as evidenced by studies [6, 7, 8, 9]. One study from the UK, conducted by Cunningham and colleagues in 2019, highlighted the superiority of shallow neural networks over regression models. Notably, this research also emphasized the potential need for distinct approaches when dealing with non-musical sounds, as their characteristics may differ from those of musical sounds. In another study, a machine learning approach incorporating audio features was introduced to predict the emotions evoked by music. This research, conducted by Er and EsİN in 2021, reported that while SVM achieved a classification success rate of 77.46%, this performance was surpassed by Artificial Neural Networks (ANN) with a success rate of 79.30% after normalizing acoustic features. Moreover, researchers from the National University of Kaohsiung proposed a novel approach called Segment-based Progressive Learning. This approach effectively combined recognition results from segments using CNN and SVM, resulting in improved recognition accuracy. In the quest to enhance personalized MER systems, a separate study investigated the integration of individual factors. The findings, as presented by Liang Xu in 2020, emphasized that individual features played a more significant role in models predicting felt emotions compared to those predicting perceived emotions. These recent research endeavors have significantly contributed to the development of Music Emotion Recognition, offering novel methodologies, insights into feature normalization, and the importance of individual factors in refining the effectiveness of MER systems.

3. Compare the CNN and SVM algorithms

In the realm of MIR, empirical evidence from prior research underscores the prominent utilization of CNNs and SVMs. While there are instances in which CNNs did not yield superior performance, it is noteworthy that the prevailing sentiment within the academic community suggests that CNNs possess untapped potential. Any subpar results can be attributed to factors such as dataset quality, training techniques, or model architectures.

CNNs offer distinct advantages within the MIR domain, notably their suitability for capturing local features such as spectral patterns and timbral characteristics. These attributes are pivotal for a spectrum of MIR tasks, including music genre classification, instrument recognition, and mood analysis. CNNs are inherently tailored to excel in tasks necessitating the detection of local patterns. Furthermore, their translation-invariance property empowers them to discern these patterns irrespective of their temporal position, a critical feature in the context of music analysis. The application of convolutional filters that traverse input data aligns seamlessly with the nature of audio signals, where informative features can manifest at various time positions. This filter-based approach empowers CNNs to efficiently glean salient insights from audio spectrograms and waveforms.

Conversely, SVMs play a significant role in MIR due to their regularization capabilities, which are crucial for handling MIR datasets characterized by limited sample sizes and extensive feature spaces. This regularization ability ensures that the models do not overfit the data, leading to more reliable and generalizable results. SVMs offer an additional advantage in the realm of feature selection. They facilitate the assessment of feature importance within the classification process, enabling the reduction of feature vector dimensionality and enhancement of overall model efficiency. In scenarios characterized by noisy data, which is an inherent characteristic of MIR datasets, SVMs

persevere by demonstrating resilience to such noise, thereby consistently delivering robust classification results.

4. Discussion

Current research trends in MIR suggest the necessity for further investigation into the potential of CNNs. While existing research has laid a foundation for the application of CNNs in MIR, a substantial portion of their capabilities remains untapped. This creates an exciting opportunity for researchers to delve deeper into CNNs' applicability within MIR, to unlock higher levels of performance and innovation. CNNs offer several advantages, including their aptitude for capturing local features, such as spectral patterns and timbral characteristics. These attributes are crucial for a wide range of MIR tasks, including music genre classification, instrument recognition, and mood analysis.

Local feature capture, an essential component of MIR, is expected to maintain its significance as a dominant trend in the field. Music comprises a complex blend of elements, and each segment of a musical composition carries a wealth of diverse features. As such, researchers are continually engaged in the pursuit of capturing and analyzing these intricate features efficiently and comprehensively. The multifaceted nature of music necessitates ongoing exploration and innovation in the methodologies employed to capture and interpret these features accurately.

Moreover, the intricate relationship between music and emotion has paved the way for advancements within the domain of Support Vector Machines (SVMs). Music is inherently characterized by its emotive components, and the ambiguity of emotional expression in music aligns with the domain of SVMs. Researchers are actively working on enhancing SVM algorithms to mitigate the impact of noise, which is a common challenge in MIR datasets. By doing so, they aim to bolster the performance of SVMs in the context of MIR, particularly in tasks where the classification of emotional content in music plays a pivotal role.

In conclusion, the future trends in MIR are marked by a ongoing exploration of CNNs, the unwavering focus on local feature capture, and the enhancement of SVM algorithms to address issues related to noise in the field. These trends collectively contribute to the ongoing evolution and maturation of Music Information Retrieval, offering a promising landscape for future research and innovation.

5. Conclusion

This paper conducts a comprehensive review of recent research within the domain of Music Information Retrieval (MIR) to present current findings and offer insights into future trends, particularly concerning the utilization of machine learning methods in MIR. MIR, characterized by its intricate nature, encompasses diverse subfields, including music genre classification, musical instrument recognition, and the notably prominent field of Music Emotion Recognition. Among these, MER stands out as the most popular and promising subtitle within the discipline. CNN and SVM algorithms hold substantial promise in shaping the future of MIR. Given that many research endeavors have yet to yield expected outcomes, a prospective trend involves delving deeper into the potential of CNN while substantiating this exploration with empirical evidence. Considering the inherent complexity of MIR and the subjectivity inherent in musical definitions, an avenue for future research involves interdisciplinary investigations. Such cross-disciplinary studies aim to enhance the precision of predictive results within this multifaceted field. In summary, the paper not only synthesizes recent research outcomes but also underlines the potential directions and imperatives for advancing MIR through the adept application of machine learning methods.

Refereces

[1] J. Y. Haojie Wei, Rui Zhang, Yueguo Chen, Gang Wang, JEPOO: Highly Accurate Joint Estimation of Pitch, Onset and Offset for Music Information Retrieval, Proceedings of the Thirty-Second International Joint Conference on Artificial Intelligence, 2023.

[2] B. Jaishankar, R. Anitha, F. Daniel Shadrach, M. Sivarathinabala, and V. Balamurugan, Music Genre Classification Using African Buffalo Optimization, Computer Systems Science and Engineering, 2023, 44 (2), 1823 - 1836.

[3] M. S. Siddharth Gururani, Alexander Lerch, AN ATTENTION MECHANISM FOR MUSICAL INSTRUMENT RECOGNITION, presented at the 20th International Society for Music Information Retrieval Conference, 2019.

[4] D. B. Prabhjyot Singh, Omkar Joshi, Nita Patil, Implementing Musical Instrument Recognition using CNN and SVM, International Research Journal of Engineering and Technology (IRJET), 2020, 6 (7); 1487 - 1493.

[5] J. Ruiz-Palmero, E. Colomo-Magaña, J. M. Ríos-Ariza, and M. Gómez-García, Big Data in Education: Perception of Training Advisors on Its Use in the Educational System, Social Sciences, 2020, 9 (4).

[6] B. M. Nicholas Farris, Richard Savery, Gil Weinberg, Musical Prosody-Driven Emotion Classification: Interpreting vocalists Portrayal of Emotions Through Machine Learning, 2021.

[7] X. W. Liang Xu, Jiaming Shi, Shutong Li, Yuhan Xiao, Qun Wan and Xiuying Qian, Effects of individual factors on perceived emotion and felt emotion of music: Based on machine learning methods, Psychology of Music,2020, 49 (5): 1069 - 1087.

[8] N. J. N. Sangeetha Rajesh, Musical instrument emotion recognition using deep recurrent neural Musical instrument emotion renceotwgnoirtkio, Procedia Computer Science, 2020, 167, 16 - 25.

[9] T.-P. H. Ja-Hwung Su, Yao-Hong Hsieh, Shu-Min Li, Effective music emotion recognition by Segment-based Progressice learning, 2020 IEEE International Conference on Systems, Man, and Cybernetics (SMC), 2020.