Analysis and Expression of Music Emotion Based on CAD and Deep Reinforcement Learning Algorithm

Ya Deng[1] and Na Lin[2]

[1]Department of Art and Sports, Huanghe Science and Technology University, Zhengzhou, Henan 457000, China, 200305010@hhstu.edu.cn
[2]Department of Art and Sports, Huanghe Science and Technology University, Zhengzhou, Henan 457000, China,13783593836@163.com

Corresponding Author: Ya Deng, 200305010@hhstu.edu.cn

Abstract. This article seeks to introduce a fresh perspective and approach to music emotion analysis by proposing a novel algorithm that integrates Deep Reinforcement Learning (DRL). The algorithm leverages the capabilities of deep neural networks to discern emotional characteristics within music autonomously. By utilizing the reinforcement learning framework, the decision-making process of the model is refined, enabling more precise identification of subtle emotional nuances in music. Experimental findings reveal that this algorithm significantly outperforms traditional methods in music emotion analysis, demonstrating a notable enhancement in detection accuracy. The key advantage of this algorithm lies in its ability to circumvent the intricacies and uncertainties associated with manual feature extraction. Furthermore, it exhibits superior adaptability to intricate music emotion analysis tasks, effectively elevating the accuracy and efficacy of classification. Following extensive training iterations, the model demonstrates a remarkable capacity to swiftly accommodate new data distributions and emotional expression patterns. In conclusion, the DRL-based music emotion analysis algorithm presented in this article contributes innovative research concepts and methodologies to the field and holds substantial theoretical and practical significance.

Keywords: CAD; Deep Reinforcement Learning; Music Emotion Analysis
DOI: https://doi.org/10.14733/cadaps.2024.S23.19-34

1 INTRODUCTION

As the core component of human culture, music has played a decisive role in our daily life since ancient times. It is not only a form of entertainment but also an important carrier of human emotional expression, communication, and exchange. From classical to modern, from east to west, the diversity of music reflects the infinite creativity of human beings for sound and also reflects people's profound demand for emotional expression. In recent years, with the development of artificial intelligence technology, the application of nonparametric fuzzy classification and evolutionary development frameworks in music sentiment analysis has gradually received attention. Abboud and Tekli [1] introduced the basic concepts of nonparametric fuzzy classification and evolutionary development frameworks, then elaborated on how to combine them for music sentiment analysis and explored their application prospects in music composition. Music sentiment analysis is an important research direction in the field of music information processing, aimed at extracting and understanding emotional content from music signals. Traditional music sentiment analysis methods are usually based on manually designed features and classifiers, but due to the complexity and uncertainty of music emotions, these methods often struggle to achieve ideal performance. It uses nonparametric fuzzy classification methods to classify the extracted emotional features of music. By introducing fuzzy mathematics theory, the uncertainty of classification boundaries can be addressed to improve the accuracy of sentiment analysis. Applying the evolutionary development framework to the optimization process of nonparametric fuzzy classifiers. By simulating the natural evolution process and iteratively searching for the optimal classifier parameters and fuzzy classification rules, the efficiency and accuracy of sentiment analysis can be improved. However, it is not easy to understand the emotions contained in music, which requires a careful analysis of musical elements such as melody, harmony and rhythm, as well as a deep understanding of cultural background and personal experience. Ashok et al. [2] proposed a facial emotion recognition method based on neural network-optimized descriptor selection. Specifically, it constructed a deep convolutional neural network (CNN) model for automatically learning feature descriptors of facial images. During the training process, it adopts an optimization strategy that combines descriptor selection and feature learning to improve the recognition performance of the network. Firstly, it utilizes pre-trained CNN models (such as VGG, ResNet, etc.) to extract low-level and high-level features of facial images. Then, a descriptor selection mechanism was introduced to construct a more compact and effective feature representation by selecting the most representative feature descriptors. The descriptor selection mechanism can be implemented using optimization algorithms such as greedy search, genetic algorithm, or reinforcement learning. To verify the effectiveness of the proposed method, we conducted experiments on publicly available facial emotion recognition datasets. The experimental results show that compared with traditional manual descriptor design methods, facial emotion recognition methods based on neural network-optimized descriptor selection have achieved significant improvements in accuracy and robustness. With the swift advancement of computer technology, the domain of artificial intelligence (AI) has witnessed unparalleled breakthroughs. Among these, deep learning, a robust machine learning technique, has demonstrated remarkable capabilities across various fields, including image recognition, speech recognition, and natural language processing. In the realm of music, deep learning harbours significant potential, particularly in the intricate task of music emotion analysis. During the process of music appreciation, people's emotional state is influenced by music, leading to changes in biophysical signals. Therefore, by analyzing biophysical signals, features related to music emotions can be extracted, providing a new basis for music emotion classification. Bălan et al. [3] introduced the research background and significance of music emotion classification, elaborated on how to use biophysical signals and machine learning techniques for music emotion classification, and explored its application prospects. During the process of music appreciation, people's emotional state is influenced by music, leading to changes in biophysical signals. Therefore, by analyzing biophysical signals, features related to music emotions can be extracted, providing a new basis for music emotion classification. Machine learning technology is a data-driven approach that can learn useful information from a large amount of data and build models that can automatically classify. In music emotion classification, machine learning techniques can utilize the features of biophysical and audio signals to construct classifiers that can automatically recognize music emotions.

Music emotion recognition is a complex task that involves extracting emotion-related features from music signals. Feature selection plays a crucial role in this process, directly affecting the accuracy and efficiency of final recognition. Chattopadhyay et al. [4] proposed a feature selection model that combines Cluster Group Generation and Balance Optimizer (CGBO) and Atomic Search Optimization Algorithm (ASOA), aiming to solve the feature selection problem in music emotion recognition. Music emotion recognition is an important research direction in the field of music information processing, with the goal of enabling machines to automatically understand and recognize emotions in music. Feature selection is a crucial step in the process of music emotion recognition, with the goal of selecting the most emotionally relevant features from the original music signal to improve recognition performance. However, due to the complexity of music signals and the subjectivity of emotions, feature selection has become a challenging task. Traditional feature selection methods often rely on manually designed features and fixed feature extraction rules, which limits the performance and generalization ability of the system. Therefore, studying automatic and efficient feature selection methods is of great significance for improving the performance of music emotion recognition. With the rapid development of artificial intelligence technology, music emotion and human emotion recognition based on learning algorithms have become a research hotspot in the field of multimedia systems. Chaturvedi et al. [5] introduced the research background and significance of music emotion and human emotion recognition. Then, the focus was on exploring the application of learning algorithms in music emotion recognition and human emotion recognition. The prospect of building a multimedia system that integrates music emotions and human emotion recognition was also discussed. Music emotion recognition refers to identifying and judging the emotions expressed in music by analyzing music signals. It combines music emotion recognition and human emotion recognition technology to build a multimedia system that integrates the functions of both, which can provide people with more intelligent and personalized services. For example, in a music recommendation system, analyzing a user's emotional state and recommending music works that match their emotions can help alleviate their stress and improve their emotional state. In the field of human-computer interaction, by identifying and understanding the emotional state of users, the system can interact with users more intelligently and improve the user experience. Methods for music emotion analysis, rooted in deep learning, typically employ deep neural networks to autonomously extract features from music. These networks are trained using extensive labelled data, enabling automatic classification and recognition of musical emotions. However, prevailing research techniques still encounter several challenges, such as effectively extracting emotion-related features from raw musical signals, addressing differences in emotional expression across various musical styles and genres, and learning effectively with limited labelled data.

To tackle these challenges, this article introduces a music emotion analysis model grounded in DRL. DRL integrates the feature extraction capabilities of deep learning with the decision-making optimization of reinforcement learning, offering a novel solution for complex learning and decision-making scenarios. In the context of music emotion analysis, the DRL model learns to refine emotional feature extraction through environmental interaction and optimizes its decision-making based on feedback, facilitating more precise emotional analysis and recognition. Music emotion recognition is one of the important applications of artificial intelligence in the field of music, which automatically determines the emotions expressed by music by analyzing music signals. However, due to the complexity and diversity of music emotions, music emotion recognition remains a challenging task. Traditional music emotion recognition methods are usually based on manually extracted audio features, and their performance is limited by the rationality and robustness of feature selection. In recent years, the rapid development of deep learning technology has provided new solutions for music emotion recognition. Dong et al. [6] proposed a deep learning model called Bidirectional Convolutional Recursive Sparse Network (BCRSN) and explored its effectiveness in music emotion recognition in detail. The BCRSN model consists of two main components: a convolutional layer and a recursive layer. The convolutional layer is responsible for extracting local features from audio signals, such as rhythm, melody, etc., while the recursive layer is responsible for capturing the temporal dependencies of these features. By combining convolutional and recursive layers, the BCRSN model can comprehensively understand the spatiotemporal characteristics of music signals. Music emotion recognition is an important research direction in the field of music information processing, aiming to extract and understand emotional content from music signals. In recent years, with the development of computer vision and deep learning technology, using spectrograms and deep vision features for music emotion recognition has become a research hotspot. Er and Aydlek [7] first introduced the research background and significance of music emotion recognition and then elaborated on how to use chromatic spectrograms and depth visual features for music emotion recognition and explored its application prospects. Traditional music emotion recognition methods are mainly based on audio signal analysis, but the expression and understanding of music emotions are not limited to audio signals; they also involve visual elements such as the expressions and actions of music performers. In recent years, with the development of computer vision and deep learning technology, using spectrograms and deep vision features for music emotion recognition has become a research hotspot.

This research holds both theoretical value and practical significance. Theoretically, the proposed DRL model contributes a fresh perspective and approach to music emotion analysis, advancing research progress in the field. Practically, accurate music emotion analysis is crucial for applications like music recommendation, music therapy, and automatic music composition. For instance, music recommendation systems can leverage users' emotional preferences and the emotional characteristics of music to recommend more emotionally resonant tracks. In music therapy, real-time monitoring of a patient's emotional states and precise matching with musical emotions can lead to more personalized therapy plans. In automatic music composition, simulating musical traits associated with different emotions can foster the creation of richer and more diverse musical pieces.

Specifically, this study introduces several innovations:

The application of the DRL framework to music emotion analysis. Unlike traditional methods that rely primarily on static deep learning models, the proposed DRL model dynamically interacts with the environment to optimize feature extraction and classification processes, enhancing the accuracy of capturing subtle musical emotions.

The design of a reinforcement learning-based strategy to guide the extraction of musical emotional features. This strategy adapts the feature extraction approach based on accurate feedback from emotion classification, enabling the model to learn to extract the most pertinent features from the original musical signal adaptively.

Addressing the scarcity of labelled data in music emotion analysis, the DRL model employs a learning algorithm with improved sample efficiency. By effectively utilizing the limited available labelled data, the model achieves superior emotion classification performance with reduced data requirements.

The article's work will proceed as follows: Firstly, a review of the relevant background and current research in music emotion analysis will be conducted, highlighting existing issues and challenges. Subsequently, the fundamental principles of DRL and its potential in music emotion analysis will be introduced. Building on this foundation, the article will elaborate on the design rationale and implementation details of the proposed DRL-based music emotion analysis model. Finally, the method's effectiveness and advantages will be validated through rigorous experiments, and its potential extensions and applications in future research will be discussed.

2 RELATED WORK

The computer-aided design system for automatic music classification plays an important role in modern music processing, retrieval, and recommendation fields. Ge et al. [8] focused on exploring how to optimize this system based on feature analysis to improve its classification accuracy and efficiency in order to meet the growing demand for music data processing. The continuous expansion of digital music libraries and the diversification of user needs have made music automatic classification technology a research hotspot in the field of music information processing. The computer-aided design system for music automatic classification based on feature analysis extracts audio features of music and combines machine learning algorithms to achieve automatic music classification. It uses feature selection algorithms to select the most representative features from a large number of extracted features and input them into the classifier, reducing feature redundancy and computational complexity and improving classification efficiency. With the continuous progress of music computer technology, its application in music education is becoming increasingly widespread. Gorbunova and Plotnikov [9] explored how music computer technology can serve as an effective tool for achieving multimodal music perception, playing its unique role in the field of education and helping students understand and experience music more comprehensively. From early audio processing software to modern virtual instruments and interactive music applications, these technologies not only enrich the forms of music creation but also bring revolutionary changes to music education. Multimodal music perception refers to the comprehensive understanding and experience of music through different perception methods, such as auditory, visual, tactile, etc. This multisensory participation not only enhances students' perception of music but also helps cultivate their creativity, collaborative ability, and critical thinking. In music classes, teachers can use interactive music software to teach students how to create and play music. In the field of music therapy, music computer technology is also used to help students with special needs improve their music perception and expression abilities.

With the development of artificial intelligence technology, the application of emotion recognition in music players is gradually receiving attention. Gupta [10] introduced the development process of a music player application based on a reinforcement learning algorithm for emotion recognition. It explores how to identify user emotions through algorithms and recommend suitable music based on them. In addition to providing music playback function, how to recommend suitable music based on user emotions has also become an important research direction. Reinforcement learning algorithms, as an adaptive decision-making method, can continuously interact with the environment to learn and find the optimal decision-making strategy. Emotion recognition technology is usually based on various information sources such as speech, text, and images, analyzing the user's expression, intonation, and text content to determine their emotional state. In a music player, users can infer their emotional state by analyzing their music playback history, song preferences, and other information. Music emotion recognition is a challenging task in the field of artificial intelligence, which involves analyzing and understanding audio signals, as well as extracting emotion-related features from them. With the development of deep learning technology, more and more research is starting to use deep neural networks for music emotion recognition. Hizlisoy et al. [11] proposed a music emotion recognition method based on a Convolutional Long Short Term Memory (ConvLSTM) deep neural network and explored its application in music emotion recognition in depth. In music emotion recognition, convolutional layers are responsible for extracting local features from audio signals, such as rhythm, melody, etc. The long-term and short-term memory layers are responsible for capturing the temporal changes of these features in order to understand the overall emotions and atmosphere of music. By combining convolutional layers and long short-term memory layers, ConvLSTM networks can more comprehensively understand music signals and improve the accuracy of music emotion recognition. It uses annotated music emotion datasets to train ConvLSTM models. Optimize the performance of the model by adjusting its parameters and structure. Traditional speech emotion recognition methods are usually based on manually designed features and classifiers, but their performance is often limited due to the complexity and uncertainty of speech signals. Huang et al. [12] proposed an application of a multi-layer hybrid fuzzy classification method based on SVM and improved PSO in speech emotion recognition. By introducing fuzzy classification and multi-layer structure, this method can better handle the uncertainty and complexity of speech signals. The improved PSO algorithm is used to optimize the parameters of SVM and the weights of fuzzy classifiers, thereby improving classification performance. The experimental results show that this method has achieved significant performance improvement on publicly available speech emotion recognition datasets. In the future, we will further study how to apply this method to other emotion-computing tasks, such as facial expression recognition, text sentiment analysis, etc., to make greater contributions to the development of the emotion-computing field.

Music emotion recognition is a popular research direction in the field of artificial intelligence, aimed at automatically analyzing and understanding emotional content in music. In recent years, convolutional neural networks (CNNs) have achieved significant results in music emotion recognition, but due to their numerous hyperparameters, effectively adjusting these parameters to improve recognition performance has become a challenge. Particle Swarm Optimization (PSO) is an optimization method based on swarm intelligence, which has the advantages of strong global search ability and fast convergence speed. Kalaiarasi [13] proposed a method of using a particle swarm optimization algorithm to optimize the hyperparameters of convolutional neural networks to improve the accuracy of music emotion recognition. It uses the training set to train the CNN model and calculates the recognition accuracy corresponding to each particle (i.e. each group of hyperparameter combinations) as the fitness value of the particles. The higher the accuracy, the better the fitness of the particles. Update the position and velocity of particles based on their fitness values and population information (such as individual optimal position and global optimal position). By continuously adjusting the position and velocity of particles, the particle swarm gradually approaches the global optimal solution. Music emotion recognition is an important research direction in the field of music information processing, with the goal of automatically identifying and understanding the emotions contained in music by analyzing music signals. Feature selection is a crucial step in the process of music emotion recognition, which is of great significance for improving recognition performance and reducing computational complexity. Li et al. [14] proposed a method for feature selection using an improved particle swarm optimization (PSO) algorithm to enhance the performance of music-based emotion recognition systems. In practical applications, extracting emotion-related features from complex music signals remains a challenge. Traditional feature selection methods often rely on manually designed features and fixed feature extraction rules, which limits the performance and generalization ability of the system. Therefore, studying automatic and efficient feature selection methods is of great significance for improving the performance of music emotion recognition. The application of online resources in music appreciation courses is gradually becoming popular. Music appreciation courses are an important way to cultivate students' aesthetic ability and cultural literacy. Traditional music appreciation courses are often limited by factors such as textbooks, teacher resources, and teaching time, making it difficult to provide rich and diverse music experiences and in-depth music analysis. The computer-aided teaching management system based on network resources can make up for these shortcomings and provide students with a more flexible and personalized learning experience. Pei and Wang [15] analyzed the design and implementation of a computer-assisted teaching management system for music appreciation courses based on network resources, exploring its advantages, challenges, and future development directions. Utilize artificial intelligence and big data technology to further improve the accuracy and efficiency of personalized recommendations and learning tracking. Strengthen cooperation with the music industry and introduce more high-quality music resources and professional guidance. Expand multi-platform support, such as mobile devices, virtual reality, etc., to provide a richer learning experience.

Rajasekhar et al. [16] proposed a new deep belief network (PSO-WO-DBN) model based on particle swarm mean updating and whale optimization to address the problem of traditional DBN models easily falling into local optima during training. This model adopts the PSO-WO hybrid optimization algorithm to optimize the parameters and structure of the DBN model in order to improve the recognition accuracy and stability of the model. Specifically, we combine the PSO-WO algorithm with the training process of the DBN model, utilizing the global search capability of the PSO-WO algorithm to find the optimal parameters and structural configuration of the DBN model, thereby preventing the model from getting stuck in local optima. To verify the effectiveness of the PSO-WO-DBN model in speech emotion recognition, we conducted experiments on publicly available speech emotion recognition datasets. The experimental results show that compared with traditional DBN models, the PSO-WO-DBN model has achieved significant improvements in recognition accuracy and stability. In addition, we also conducted a detailed analysis of the performance of the PSO-WO optimization algorithm, verifying its effectiveness in improving the performance of the DBN model. Music, as a medium of emotional expression, can directly touch people's hearts. Different types of music, such as classical, pop, rock, etc., often convey different emotions. The research on music emotion recognition aims to extract and understand emotional content from music signals, providing strong support for music recommendation, music creation, and other fields. In recent years, machine learning technology has played an important role in music emotion recognition, injecting new vitality into the development of this field. Machine learning technology plays an important role in music emotion recognition. By training a large amount of music data, machine learning models can learn the mapping relationship between music and emotions, thereby achieving automatic recognition of music emotions. In order to verify and compare the performance of different machine learning algorithms in music-type emotion recognition, Ramírez and Flores [17] selected various types of music data, including classical, pop, rock, etc. Then, it uses different machine learning algorithms to train these music data and compares their emotion recognition performance.

Music emotion recognition is one of the core tasks in the field of music information processing, with the goal of automatically recognizing and understanding the emotions contained in music. In recent years, deep learning, especially recurrent neural networks (RNNs) and their variants, such as long short-term memory networks (LSTM), have been widely used for music emotion recognition. However, selecting appropriate features and hyperparameters is crucial for the performance of LSTM. To this end, Tripathi and Choudhury [18] proposed a music emotion recognition method that combines LSTM and an improved Mouse Swarm Optimization Algorithm (IMOA), aiming to automatically optimize the parameters and feature selection of the LSTM model, thereby improving recognition accuracy. Music emotion recognition is an important research direction in the field of music information processing, which aims to identify and understand the emotions contained in music by analyzing music signals. LSTM, through its unique gating mechanism and memory unit, can capture long-term dependencies in music signals, effectively extracting emotion-related features. Traditional feature selection methods often rely on manually designed features and fixed feature extraction rules, which limits the performance and generalization ability of the system. Meanwhile, the selection of hyperparameters often relies on empirical or heuristic methods, lacking systematic and global optimization capabilities. Music, as a universal art form, can trigger and convey various emotions. However, different individuals may have different emotions and feelings towards music. Xu et al. [19] explored how individual factors affect music perception emotions and sensory emotions and analyzed them based on machine learning methods. By collecting a large amount of data and using advanced machine learning algorithms, we analyzed the effects of factors such as age, gender, cultural background, and music training on music emotions and feelings. Music is an art form that transcends cultural and linguistic boundaries, capable of inspiring and conveying rich emotions. Different types and styles of music can trigger different emotional responses from the audience. However, these reactions are influenced not only by the music itself but also by individual factors. Understanding how these factors affect music perception and emotional perception is crucial for music creation, performance, and appreciation.

3 MUSIC EMOTION ANALYSIS MODEL BASED ON DRL

As computer science and technology continue to advance, the utilization of computer-aided design (CAD) has steadily broadened beyond its conventional engineering and architectural domains, encompassing a more extensive array of cultural and creative industries. Although the application of CAD technology in the music field is not as obvious as that in graphic design or mechanical manufacturing, its unique value and great potential in music creation, analysis and communication can not be ignored. Especially in music emotion analysis, CAD technology plays a vital role. In addition, CAD technology also plays an important role in constructing a musical emotion analysis model. Traditional musical emotion analysis often relies on the auditory experience and subjective feeling of the analyst, but this method has the problems of strong subjectivity and poor consistency. By utilizing CAD technology, we are able to construct a precise mathematical model rooted in musical components and emotional tags, effectively converting emotional analysis into a quantifiable and calculable mathematical formulation. This innovative approach not only enhances the precision and objectivity of sentiment analysis but also establishes a robust groundwork for future automated analytical processes.

The use of CAD in music emotion analysis does not exist in isolation. In order to give full play to its potential, we need to integrate it with cutting-edge machine learning technologies such as deep learning and reinforcement learning. By automatically extracting music features through the deep neural network, optimizing the process of feature extraction and classification by reinforcement learning, and combining the visualization and model-building capabilities of CAD technology, we can build a more accurate and efficient music emotion analysis system. Such a system will not only promote the in-depth development of academic research but also be widely used in practical scenes such as music recommendation, music therapy and automatic music creation, bringing more colourful music experiences to people's lives. As AI technology continues to evolve, Deep Reinforcement Learning (DRL), a sophisticated machine learning approach that integrates the strengths of deep learning and reinforcement learning, has gradually demonstrated its distinctive benefits and potential in the realm of music emotion analysis. This model excels not only in uncovering latent features within music but also in precisely enhancing the efficiency of emotion classification through a reinforced learning mechanism. Consequently, it offers a fresh perspective and solution for research and applications in music emotion analysis.

DRL stands out as an ingenious machine learning technique that adeptly merges the feature extraction capabilities of deep neural networks with the decision-making optimization of reinforcement learning algorithms. In this framework, the deep neural network takes on the task of extracting meaningful features from raw data, while the reinforcement learning algorithm refines the model based on these features, enabling it to make optimal decisions. Within the context of music emotion analysis, the DRL model is capable of independently recognizing musical features closely tied to emotions by processing large volumes of labelled music data, thereby continuously refining the precision of emotion classification.

The framework of music emotion classification and recognition based on DRL is a comprehensive system that starts from the music database and brings together many music samples with different styles and rich emotions (Figure 1). At the beginning of system operation, preprocessing plays a key role, which is responsible for cleaning the original music data, ensuring the consistency and quality of input information, and then laying a solid foundation for subsequent feature extraction. Emotional feature extraction is the core link in this framework, where the deep neural network makes great efforts to learn and mine the features closely related to emotion from the preprocessed music signals, such as the ups and downs of melody and the level of harmony. These features then constitute a data set, which becomes an important basis for training and verifying the emotion classification model. With the help of DRL, the classifier or regressor can be continuously optimized, and its output predicted labels are compared with the real labels marked by human beings in the database so as to accurately evaluate the model performance. In the end, this framework not only realizes the in-depth analysis and understanding of musical emotions but also expresses these emotions in many ways, whether by generating emotional descriptions, recommending music with similar emotions, or using visual tools to show the emotional context of music, which provides users with a richer and more immersive musical experience.

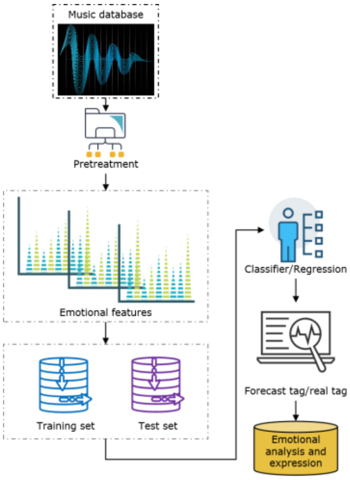

Figure 1: Framework for emotion analysis and recognition.

Training the DRL-based music emotion analysis model needs a large number of labelled music samples as support. These samples should cover music fragments with various styles and emotions and be equipped with corresponding emotional labels. Through the utilization of these samples for training, the model gains the ability to comprehend the intricate relationship between musical features and emotional tags. In the initial stages of training, supervised learning techniques can be employed to pre-train the deep neural network, enabling it to extract fundamental musical features. Subsequently, the reinforcement learning algorithm comes into play, fine-tuning and optimizing these features to elevate the precision of emotion classification. Simultaneously, by consistently interacting with the environment and learning model, the system gradually adapts to a diverse array of complex musical and emotional scenarios, thereby enhancing the stability and robustness of its classification capabilities. The DRL model tailored for music emotion analysis is visually represented in Figure 2.

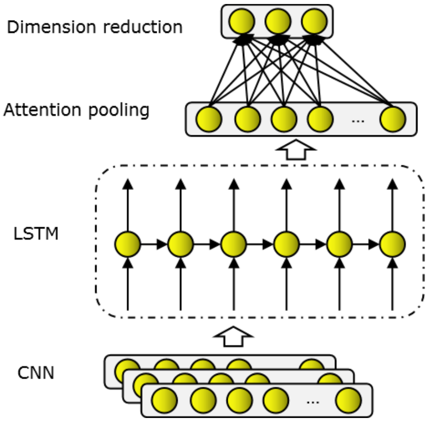

Figure 2: DRL model of music emotion analysis.

By multiplying the initial music signal, denoted as $s(n)$, with a sliding window function, denoted as $w(n)$, which has a defined window length, we obtain a windowed music signal, represented as $s_w(n)$:

$$s_w(n) = s(n) * w(n) \tag{1}$$

The application of a window function to the music signal serves to mitigate spectrum leakage resulting from framing. Two widely adopted window functions exist: the Hamming window and the rectangular window. Their respective functional expressions are provided below (The length of the frame is represented by n).

Rectangular window:

$$w(n) = \begin{cases} 1, 0 \leq n \leq (N-1) \\ 0, n = else \end{cases} \tag{2}$$

Hamming window:

$$w(n) = \begin{cases} 0.54 - 0.46\cos\left[2\pi n / (N-1)\right], 0 \leq n \leq (N-1) \\ 0, n = else \end{cases} \tag{3}$$

Selecting an optimal window function enhances the short-term parameters' ability to accurately capture the unique features of music signals while minimizing the impact of the window function itself on the analytical parameters of these short-term signals.

"Information entropy" represents the average quantity of information remaining after the elimination of redundancies, thereby signifying the level of uncertainty or disorder within the information. The mathematical expression for determining information entropy is as follows:

$$I\ \ S_1,S_2,S_3,\ldots,S_m\ \ = -\sum\nolimits_{i=1}^{m} P_i \log_2\ P_i \tag{4}$$

Within the given context, S represents a sample set with m distinct classes labeled as $C_i\ \ i=1,2,3,\ldots,m$. S_i Denotes the count of samples belonging to a particular class C_i , while P_i signifies the likelihood of any sample being part of C_i , specified as $S_i\,/\,S$. To determine the most informative sample classification attributes, we calculate the information gain post-branching, which is outlined as follows:

$$Gain\ \ A\ = I\ \ S_1,S_2,S_3,\ldots,S_m\ \ - E\ \ A \tag{5}$$

In this context, $I\ \ S_1,S_2,S_3,\ldots,S_m$ it represents the anticipated information entropy associated with a specific sample denoted as S . The Equation for determining $E\ \ A$, which represents a relevant metric is outlined below:

$$E\ \ A\ = -\sum\nolimits_{j=1}^{v} \frac{S_{1j}+S_{2j}+S_{3j}+\ldots+S_{mj}}{s} I\ \ S_{1j}+S_{2j}+S_{3j}+\ldots+S_{mj} \tag{6}$$

Information gain reflects the change in variable values that occurs when a dataset is segmented into smaller subsets. To mitigate this variance, we employ the Equation for the information gain ratio:

$$SplitInfo\ \ S,v\ = \sum\nolimits_{i=1}^{m} \frac{|S_i|}{|S|} \times \log_2 \frac{|S_i|}{|S|} \tag{7}$$

To obtain the gain ratio, one can utilize the Equation (8):

$$GrainRatio = \frac{Grain\ \ S,v}{SplitInfo\ \ S,v} \tag{8}$$

For the music segment T_i , upon completion of preprocessing, the sentence is first scanned for positive emotional words using the emotional dictionary as a reference, with an initial emotional score of $posSScore_0 = 1$ assigned. Subsequently, leveraging the results of semantic dependency analysis, a judgment is made regarding whether the identified words are qualified by negative modifiers or degree adverbs. Based on this assessment, the emotional score is adjusted in accordance with Equation (9):

$$posSScore = \prod_{i=1}^{m} \beta \prod_{j=1}^{n}\ -1\ \times posSScore \tag{9}$$

Within this context, β it represents the magnitude value assigned to degree adverbs. Based on the preceding analysis, we are able to derive the weighted value of emotional sentences containing positive emotional words within each segment:

$$posSScore = \frac{1}{l}\sum_{i=1}^{l} posSScore_i \tag{10}$$

With $posSScore_i > 0$ set to 1, it is observed that Equation (10) facilitates the normalization of eigenvalues, confining their range to [0,2]. Hence, under this characteristic, the sample can adopt L eigenvalues as prescribed by Equation (11):

$$L =\ \max - \min\ \ / \, b \tag{11}$$

In this context, $\max$ they $\min$ represent the upper and lower bounds of the interval (whether open or closed), respectively. b Denotes the step size, which can be set to 0.5. Based on these parameters, the positive emotional word features of the sample can adopt four distinct eigenvalue values.

4 RESULT ANALYSIS AND DISCUSSION

4.1 Experimental Environment

The experimental environment is the key factor to ensure the accuracy and repeatability of this study. In this experiment, computer hardware equipped with high-performance processors, large-capacity memory and storage devices is selected to support the training and reasoning needs of the deep learning model. At the same time, a stable Linux operating system is adopted, and deep learning frameworks such as TensorFlow and Keras, as well as related dependency libraries and toolkits, are installed to build a perfect software environment. In order to analyze music emotion, we use public data sets and audio processing libraries such as Librosa to process music signals and extract features. During the whole experiment, the operation was strictly in accordance with the experimental design to ensure the consistency and comparability of the experiment, and the experimental environment was fully tested and verified to ensure its stability and reliability. This experimental environment provides solid support and guarantee for the task of music emotion analysis and ensures that we can obtain accurate and credible experimental results.

4.2 Result Analysis

The results show the change in the loss-fitting curve of the model in the initial and final training stages. Comparing Figure 3 and Figure 4, we can perceive the optimization and convergence of the model in the training process. By comparing the changes in the loss fitting curve between the first training and the final training, we can deeply analyze the optimization and convergence of the model in the training process, as well as the model performance and improvement space reflected by these data.

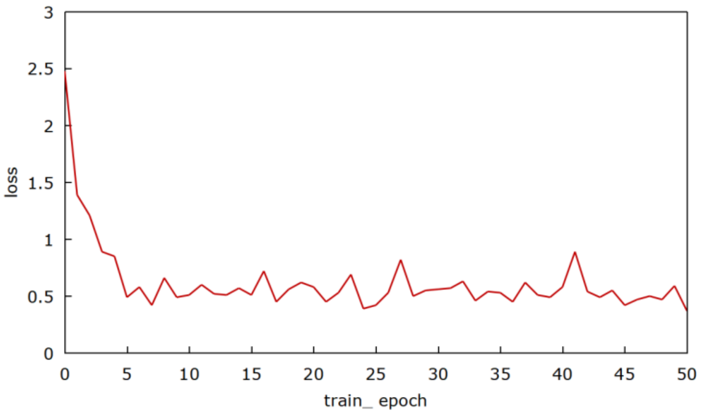

Figure 3: First training loss curve.

In the first training, the model is faced with brand-new data sets and tasks, so the initial loss is large. At this stage, the model is trying to adapt to the distribution and characteristics of data and reduce the prediction error by constantly adjusting parameters. With the increase of training rounds, we can see that the loss value decreases gradually, which shows that the model is gradually learning the inherent laws of data and the mode of emotional expression.

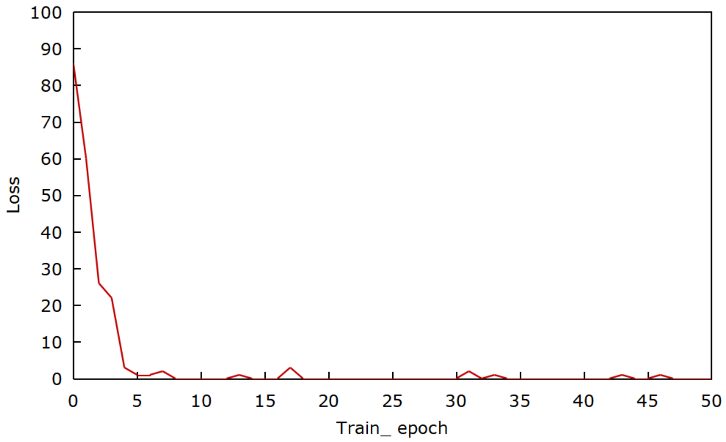

Figure 4: Loss curve of 20th training.

After the ninth round, the loss curve tends to be stable, which means that the model has basically mastered the characteristics of the data and reached a relatively good performance level. In the final training, the performance of the model has been further improved. After 20 rounds of iterative training, the model reached a steady loss in the fourth round. This shows that with the training, the model has a deeper understanding of data characteristics and can adjust parameters to adapt to new data more quickly.

To provide a thorough evaluation of the DRL music emotion analysis algorithm introduced in this article, we have conducted a comparison against the conventional method of music emotion analysis. By referring to the data presented in Table 1, it becomes evident that our proposed algorithm offers significant advantages in terms of accuracy when it comes to music emotion analysis.

Music number	*This method*	*RNN*	*SVM*
1	97.475	83.316	90.879
2	96.351	82.527	88.625
3	93.294	82.213	90.911

Table 1: Accuracy of music emotion analysis by different methods.

The algorithm's accuracy stands at 95.71%, notably surpassing the two traditional methods. This underscores the superior recognition precision and stronger generalization capabilities of the DRL model introduced in this article for music emotion analysis tasks. This advantage stems from the DRL model's ability to automatically extract emotional features from music and refine its decision-making through environmental interaction, enabling a more precise capture of musical emotions. In contrast, traditional approaches often hinge on manually crafted features and static classifiers for music emotion analysis. Our algorithm, however, leverages deep neural networks for automatic feature extraction and reinforces learning for model parameter optimization. This end-to-end training approach enhances the model's adaptability to the intricacies of music emotion analysis, thereby elevating its performance.

To delve deeper into the expressive capabilities, classification performance of the three descriptors, and the impact of training sample size on detection accuracy, this study conducted detection accuracy tests across varying offline training sample sets. A comparative analysis of the experimental results presented in Figures 5, 6, and 7 yields the following insights.

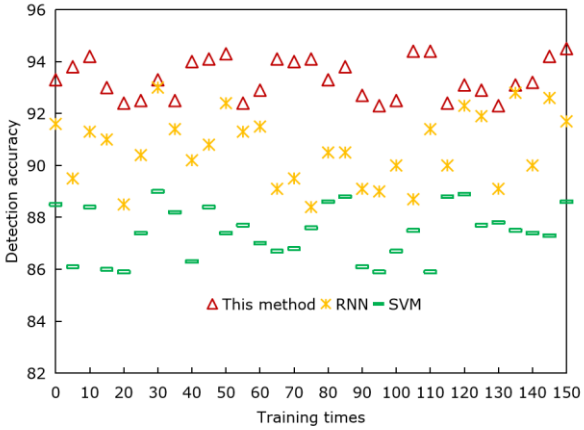

Figure 5: Music 2 classification performance.

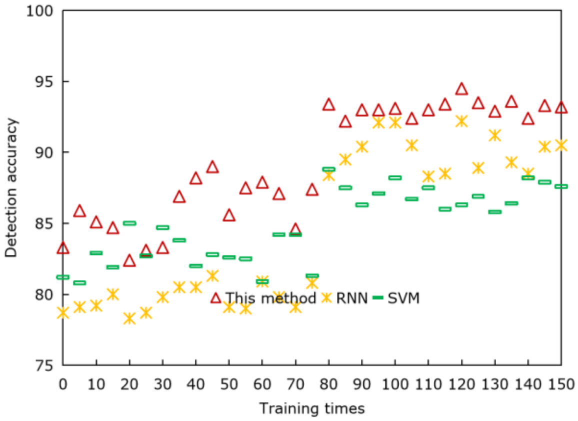

Figure 6: Music 3 classification performance.

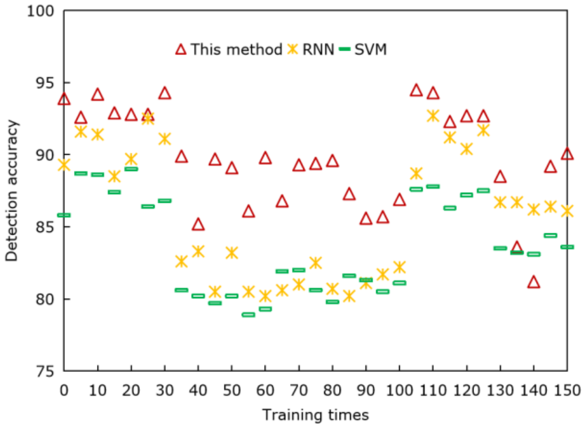

Figure 7: Music 4 classification performance.

Upon comparing the detection accuracy of the three methods, it becomes evident that the algorithm presented in this article consistently attains high accuracy across varying training sample sets. This underscores the algorithm's efficacy and excellence in music emotion analysis. Through the integration of deep learning and reinforcement learning technologies, our algorithm adeptly extracts emotional features from music and classifies them with precision. This seamless, end-to-end training approach allows the model to intricately grasp the subtleties of musical emotions, thereby elevating its classification proficiency. Notably, even with a limited number of training samples, our algorithm maintains impressive detection accuracy. This highlights its efficient utilization of training data and its ability to discern effective emotional patterns within constrained datasets.

4.3 Application and Prospect of Model

The DRL-based music emotion analysis model offers promising applications across diverse domains such as music recommendation, music therapy, and automatic music composition. In music recommendation systems, this model can delve into users' musical preferences and emotional needs, curating a playlist that aligns with their current emotional state. In the realm of music therapy, it can tailor individualized programs by gauging patients' emotional states. For automatic music creation, this model can compose pieces with targeted emotional expressions.

As technology continues to evolve, the DRL-based model is poised to usher in a more intelligent and automated era of music emotion analysis. Real-time analysis of musical emotions, for instance, can enhance the interactivity and intelligence of musical performances. Furthermore, integrating this model with other AI technologies can lead to the development of a comprehensive and sophisticated musical emotion analysis system. In summary, the DRL-based music emotion analysis model ushers in innovative approaches and methodologies for exploring and applying music emotion analysis, with vast potential in multiple fields. With the advent of new technologies, this model is destined to play a pivotal role in the future of musical emotion analysis.

5 CONCLUSION

Music emotion analysis, a pivotal intersection of music information retrieval and artificial intelligence, has garnered significant attention. This article introduces a groundbreaking DRL-powered algorithm for music emotion analysis, showcasing its distinct advantages over conventional methods. The findings underscore the algorithm's proficiency in capturing subtle musical emotions, marking a paradigm shift in music emotion analysis.

The algorithm leverages the deep neural network to seamlessly extract emotional features from music, eliminating the reliance on manual feature design prevalent in traditional methods. Reinforcement learning refines the model's decision-making based on real-time feedback, ensuring precision in motion capture. This end-to-end approach imbues the model with remarkable adaptability, maintaining robust classification performance even in complex scenarios.

Following iterative training, the model demonstrates swift adaptability. Whether confronted with shifting data distributions or varying emotional expressions, it rapidly adjusts parameters to acclimate to new environments. This trait is crucial for practical deployments, ensuring the model remains resilient to evolving data and task demands. Compared to traditional counterparts, our algorithm excels in detection accuracy across diverse offline training samples, attesting to its efficient data utilization and ability to discern effective emotional patterns within limited datasets.

Future endeavours will focus on optimizing the model's architecture, enhancing training efficiency, and tackling more intricate music emotion analysis challenges. These efforts aim to propel the advancement and application of music emotion analysis technology.

Ya Deng, https://orcid.org/0009-0005-5756-8267
Na Lin, https://orcid.org/0009-0008-9768-4037

REFERENCES

[1] Abboud, R.; Tekli, J.: Integration of nonparametric fuzzy classification with an evolutionary-developmental framework to perform music sentiment-based analysis and composition, Soft Computing, 24(13), 2020, 9875-9925. https://doi.org/10.1007/s00500-019-04503-4

[2] Ashok, K.-P.-M.; Maddala, J.-B.; Martin, S.-K.: Enhanced facial emotion recognition by optimal descriptor selection with neural network, IETE Journal of Research, 69(5), 2023, 2595-2614. https://doi.org/10.1080/03772063.2021.1902868

[3] Bălan, O.; Moise, G.; Petrescu, L.; Moldoveanu, A.; Leordeanu, M.; Moldoveanu, F.: Emotion classification based on biophysical signals and machine learning techniques, Symmetry, 12(1), 2019, 21. https://doi.org/10.3390/sym12010021

[4] Chattopadhyay, S.; Dey, A.; Singh, P.-K.; Ahmadian, A.; Sarkar, R.: A feature selection model for speech emotion recognition using clustering-based population generation with hybrid of equilibrium optimizer and atom search optimization algorithm, Multimedia Tools and Applications, 82(7), 2023, 9693-9726. https://doi.org/10.1007/s11042-021-11839-3

[5] Chaturvedi, V.; Kaur, A.-B.; Varshney, V.; Garg, A.; Chhabra, G.-S.; Kumar, M.: Music mood and human emotion recognition based on physiological signals: a systematic review, Multimedia Systems, 28(1), 2022, 21-44. https://doi.org/10.1007/s00530-021-00786-6

[6] Dong, Y.; Yang, X.; Zhao, X.: Bidirectional convolutional recurrent sparse network (BCRSN): an efficient model for music emotion recognition, IEEE Transactions on Multimedia, 21(12), 2019, 3150-3163. https://doi.org/10.1109/TMM.2019.2918739

[7] Er, M.-B.; Aydilek, I.-B.: Music emotion recognition by using chroma spectrogram and deep visual features, International Journal of Computational Intelligence Systems, 12(2), 2019, 1622-1634. https://doi.org/10.2991/ijcis.d.191216.001

[8] Ge, M.; Tian, Y.; Ge, Y.: Optimization of computer aided design system for music automatic classification based on feature analysis, Computer-Aided Design and Applications, 19(S3), 2021, 153-163. https://doi.org/10.14733/cadaps.2022.S3.153-163

[9] Gorbunova, I.-B.; Plotnikov, K.-Y.: Music computer technologies in education as a tool for implementing the polymodality of musical perception, Musical Art and Education, 8(1), 2020, 25-40. https://doi.org/10.31862/2309-1428-2020-8-1-25-40

[10] Gupta, K.-S.: Development of music player application using emotion recognition, International Journal for Modern Trends in Science and Technology, 7(1), 2021, 54-57. https://doi.org/10.46501/IJMTST070112

[11] Hizlisoy, S.; Yildirim, S.; Tufekci, Z.: Music emotion recognition using convolutional long short term memory deep neural networks, Engineering Science and Technology an International Journal, 24(3), 2020, 760-767. https://doi.org/10.1016/j.jestch.2020.10.009

[12] Huang, S.; Dang, H.; Jiang, R.; Hao, Y.; Xue, C.; Gu, W.: Multi-layer hybrid fuzzy classification based on SVM and improved PSO for speech emotion recognition, Electronics, 10(23), 2021, 2891. https://doi.org/10.3390/electronics10232891

[13] Kalaiarasi, P.: Optimizing convolutional neural network using particle swarm optimization for face recognition, Turkish Journal of Computer and Mathematics Education (TURCOMAT), 12(11), 2021, 3672-3679. https://doi.org/10.17762/turcomat.v12i11.6450

[14] Li, Z.; Qiu, L.; Li, R.; He, Z.; Xiao, J.; Liang, Y.; Pan, J.: Enhancing BCI-based emotion recognition using an improved particle swarm optimization for feature selection, Sensors, 20(11), 2020, 3028. https://doi.org/10.3390/s20113028

[15] Pei, Z.; Wang, Y.: Analysis of computer aided teaching management system for music appreciation course based on network resources, Computer-Aided Design and Applications, 19(S1), 2021, 1-11. https://doi.org/10.14733/cadaps.2022.S1.1-11

[16] Rajasekhar, B.; Kamaraju, M.; Sumalatha, V.: A novel speech emotion recognition model using mean update of particle swarm and whale optimization-based deep belief network, Data Technologies and Applications, 54(3), 2020, 297-322. https://doi.org/10.1108/DTA-07-2019-0120

[17] Ramírez, J.; Flores, M.-J.: Machine learning for music genre: multifaceted review and experimentation with audioset, Journal of Intelligent Information Systems, 55(3), 2020, 469-499. https://doi.org/10.1007/s10844-019-00582-9
[18] Tripathi, A.; Choudhury, T.: EEG based emotion recognition using long short term memory network with improved rat swarm optimization algorithm, Revue d'Intelligence Artificielle, 37(2), 2023, 281-289. https://doi.org/10.18280/ria.370205
[19] Xu, L.; Wen, X.; Shi, J.; Li, S.; Xiao, Y.; Wan, Q.; Qian, X.: Effects of individual factors on perceived emotion and felt emotion of music: based on machine learning methods, Psychology of Music, 49(5), 2021, 1069-1087. https://doi.org/10.1177/0305735620928422