International Journal of Education, Culture and Society
2023; 8(6): 261-267
http://www.sciencepublishinggroup.com/j/ijecs
doi: 10.11648/j.ijecs.20230806.16
ISSN: 2575-3460 (Print); ISSN: 2575-3363 (Online)

Analysis of the Similarity Estimation Schemes for Music and Applications

Yanjun Chen, Ning Li*

Department of Music, Shenzhen University, Shenzhen, China

Email address:
ron916@naver.com (Ning Li)

*Corresponding author

To cite this article:
Yanjun Chen, Ning Li. (2023). Analysis of the Similarity Estimation Schemes for Music and Applications. *International Journal of Education, Culture and Society*, *8*(6), 261-267. https://doi.org/10.11648/j.ijecs.20230806.16

Received: November 6, 2023; **Accepted**: November 28, 2023; **Published**: November 29, 2023

Abstract: With the maturation of big data technology and artificial intelligence algorithms within music information retrieval, this research delves into the nuanced landscape of music similarity computation and evaluation methods, along with their multifaceted applications. Positioned within the broader music information retrieval domain, the study addresses pivotal challenges utilizing advanced technologies. Central to the investigation is the exploration of music similarity detection, a vital facet of music information retrieval crucial for tasks like music plagiarism identification, song classification, and the development of music recommendation systems. The study meticulously introduces various applications of similarity computation and meticulously dissects the principles and processes of music feature extraction, incorporating methodologies such as mel-frequency cepstral coefficients, harmonic pitch class profile, convolutional neural networks, and recurrent neural networks. A comprehensive survey of prevailing models and approaches for computing similarity is presented. Beyond conventional measures like cosine and Euclidean distances, the research scrutinizes the integration of artificial intelligence algorithms and models, notably support vector machines, into the computation of music similarity. The study meticulously outlines the advantages and limitations of these methodologies, offering nuanced insights. These findings serve as a valuable reference for researchers aiming to comprehend the intricacies of music similarity computation, providing a foundation for refining existing models. The study accentuates the synergy between big data, artificial intelligence, and music information retrieval, envisaging a landscape where these technologies collectively propel the field forward.

Keywords: Music Similarity, Music Information Retrieval, Artificial Intelligence

1. Introduction

In late 1945, computers were born in the United States, and some early pioneers in electronic music exploration recognized that this new device would produce marvel-ous music. For example, in 1957, Lejaren used computer software and hardware to generate the first piece of music titled "the sliver scale." That same year, Hiller com-posed the first truly algorithmic music composition called the "Iliac Suite." First, specialized computer music research institutions were established in universities in the United States and Western Europe. Then, in 1975, the International Computer Music Association was founded, which also published the "Computer Music" magazine. As digital technology and the era of digital information technology advanced in the 1980s, computer music rapidly developed. Music compositions could be stored and disseminated digitally, leading to the gradual development of digital music. People began to use algorithmic programs to assist in computer music composition, giving rise to software like AIVA used for musical composition and orchestration. Additionally, computers were used to control analog synthesizers, creating digital analog music computers. The digitization technology in computer music was also applied to music production and performance [1, 2].

The field of MIR (Music Information Retrieval) includes song information processing, rhythm detection, harmony information retrieval, melody detection and ex-traction, music search, high-level semantic analysis, intelligent composition, and other audio processing [3]. The music search mentioned encompasses music version recognition or cover song identification, humming and singing retrieval, as well as music genre classification, which helps us gain a

better understanding of most relevant technologies [4]. Research on music similarity is a content-based music information retrieval in the field of music information retrieval. It can be categorized into long segment similarity and short segment similarity based on the length of matching fragments. For example, in the tasks of music cover detection or multi-version music recognition, there are longer similar segments, while in tasks like humming retrieval or music borrowing, there are shorter similar segments [3].

In the study of music content similarity comparison, researchers commonly extract pitch class profile (PCP) as the most representative audio feature, which effectively represents the melodic characteristics of music. Based on this, researchers proposed various variant features, such as PCP based on instantaneous frequency (IF), melodic PCP (MPCP) based on the auditory characteristics of the human ear, IFPCP to de-scribe the melodic features of music, and MFCC to supplement the low-frequency features of music. However, studies have shown that the fusion of multiple feature information can better describe music. Therefore, it combines IFPCP features and MFCC features to study music similarity. A new similarity fusion model was developed using the similarity network fusion (SNF) technique based on Qmax and Dmax [5]. The experimental results demonstrated that this model had a competitive recognition accuracy and classification accuracy compared to existing fusion methods at that time. A new model for music similarity comparison called CDmax_SVM is proposed, which combines Dmax, cosine distance, and SVM. This model achieves higher accuracy and test rates compared to both Dmax_SVM and CD_SVM models. With the increasing maturity of machine learning algorithms, there are also more methods for comparing music similarity. This article will present specific approaches and processes for evaluating similarity.

This article mainly summarizes how to judge similarity in previous music similarity evaluation methods, including the principles of feature extraction related to similarity and models for calculating similarity, etc. It provides references for more scholars who want to understand music similarity and for scholars who want to innovate better models. It also proposes a general direction for efforts to be made in the development of music similarity detection. The rest part of the paper is organized as follows. The Sec. 2 mainly introduces the principles and process of feature extraction related to similarity. The Sec. 3 introduces the main models and schemes for calculating similarity. The Sec. 4 mainly introduces the application of calculating similarity. The Sec. 5 discusses the limitations of current music similarity calculation and applications. The Sec. 6 concludes the article by summarizing the main findings, identifying limitations discovered by the author, and providing prospects for future work.

2. Basic Descriptions of Similarity for Music

Music similarity is an important concept in music information retrieval, and the concept of music similarity is very complex because it encompasses many aspects. Previously, the definition of music similarity is considered to be ambiguous and unclear because each individual's judgment and perception of music similarity can vary [6]. Some study also views it as highly subjective [7], while others suggests that music similarity has different meanings in different contexts [8]. Another study also considers music similarity to be a completely subjective concept [9]. Therefore, it can be observed that music similarity does not have the judgment of music similarity can be divided into subjective and objective aspects. Subjective judgment of music similarity includes factors such as melody, harmony, rhythm, tempo, timbre, style, genre, and emotion. On the other hand, objective judgment involves analyzing music features using signal processing and machine learning techniques. This indicates that there are two different approaches to assessing the similarity of music: one is by collecting metadata about the music, and the other is by analyzing the music signals themselves [6]. Subjectively, one can judge it to sound similar based on our intuition, but one cannot be certain if it is indeed similar. It requires further analysis of the musical signals to make a determination.

One study employs context-based similarity estimation techniques, which can be categorized into three main areas [10]: first, methods based on text retrieval, which involve searching for lyrics, writing tags, etc.; second, methods based on co-occurrence; and third, the analysis of user listening habits using collaborative filtering methods. By analyzing common preferences and behavioral patterns among user groups, the collaborative filtering algorithm is used to infer the similarity of music. For example, if a group of users frequently saves or plays similar music, it is likely that these songs are similar in certain aspects. A method that utilizes fused block-level features for music similarity estimation was used [6]. They proposed a novel approach and introduced three new block-level audio features. Previous studies employed similarity network fusion techniques to improve performance in terms of recognition ac-curacy and classification efficiency among different music versions [5, 11]. Other researches combined Dmax, cosine distance, and SVM into a new model called CDmax_SVM for comparing the similarity of music segments [3]. Several techniques for melody similarity matching were mentioned, including distance-based similarity, quantization-based similarity, and fuzzy-based similarity [4]. The estimation method for music similarity is based on the generation of overall similarity matrix using block-level feature similarity and label affinity-based similarity [12]. The discrimination of similarity can be divided into distance and similarity measures, such as Euclidean distance and Manhattan distance. One common similarity measure is cosine similarity. There are also artificial intelligence techniques used to calculate the similarity of music.

3. Feature Extraction

3.1. CPCP

Music feature extraction was previously used in chord recognition [13], using a meth-od based on Pitch Class Profile (PCP) which maps a certain frequency range of short-time spectra through techniques such as STFT and FFT to 12 pitch chroma or classes [14]. In recent years, the Critical Pitch Class Profile (CPCP) method has been pro-posed by introducing the concepts of equal loudness contour and critical band [15]. Firstly, the preprocessed audio is segmented into frames and compensated for the frequency characteristics related to the structure of the human ear, as well as the frequency selection characteristics, through loudness contour filtering and auditory filter bank filtering. The auditory representation is obtained by half-wave rectifying and down-sampling. These data are then sampled by PCP and dimensionally reduced through NMF to obtain 12 different music features. This method takes into account the relevant transmission characteristics of the outer ear structure, the frequency selection characteristics of the cochlea, and the response of hair cells to the output, combining them with the PCP method to obtain the final features.

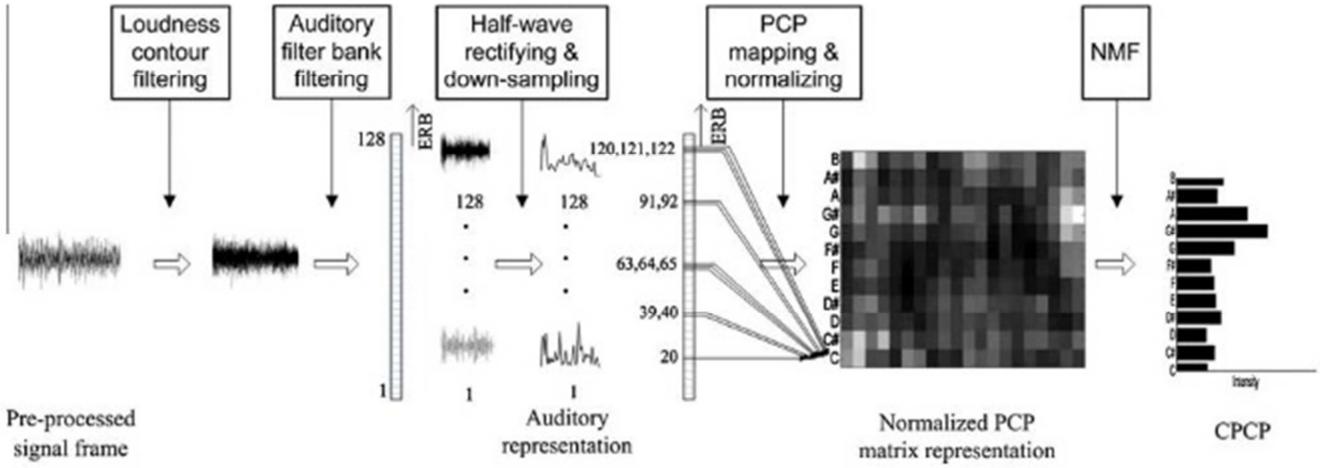

Figure 1. *Using HPCP for feature extraction, SNF is a fusion algorithm used to extract different music features [11].*

3.2. HPCP+SNF

HPCP (harmonic PCP), as a stronger version of PCP, detects the main harmonics in each frequency band and calculates their energy or amplitude. The energy or amplitude of each normalized pitch class, classified by the same pitch as PCP, is then sequentially connected to form the final HPCP feature vector [16]. Subsequent researchers have combined HPCP with similarity network fusion (SNF) [5] to identify songs and cover versions. This feature extraction method extracts tracks using HPCP, obtains similarity matrices through Dmax and Qmax matrices respectively, and finally fuses them using SNF to obtain the fused similarity matrix, as shown in Figure 2. The key idea behind this method is to obtain different features by using different similarity matrices for the same descriptor.

3.3. Others

In conclusion, the principle of feature extraction is to use mathematical methods, such as PCP, MFCC, etc., to transform the time-domain audio signal into a spectrogram with feature information, and then perform feature extraction. The feature extraction process of music similarity is generally as follows: First, the audio data needs to be preprocessed by converting the audio file into a digital audio signal. Then, methods such as Fourier transform (STFT) are used to convert the audio signal into a spectrogram. Next, features are extracted from the spectrogram, such as Mel-Frequency Cepstral Coefficients (MFCC). After that, feature selection is performed to choose the appropriate features. Finally, the extracted features are normalized, etc. Finally, similarity calculation is performed to calculate the similarity between music features, common methods include cosine similarity and Euclidean distance. If machine learning methods are used, there will be additional methods and steps. In terms of methods, convolutional neural networks (CNN) can be used for feature extraction in the time-frequency domain, and recurrent neural networks (RNN) can be used for feature ex-traction in the time domain. In terms of steps, there is an additional step of training the model using algorithms, continuously optimizing the model parameters to improve computational accuracy, and evaluating the model, such as existing models like Sup-port Vector Machines (SVM).

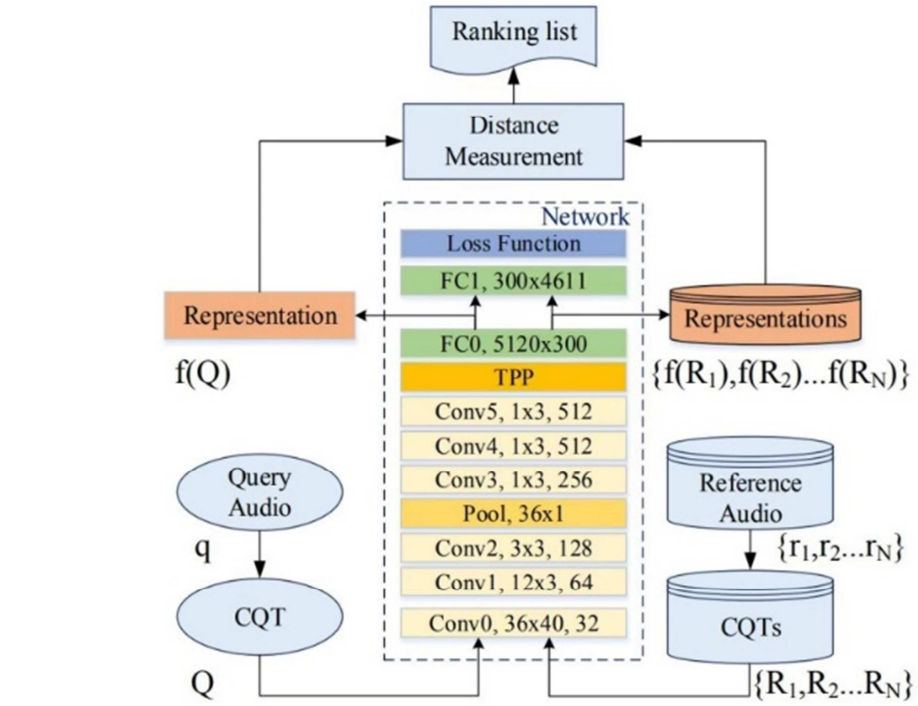

Figure 2. *Music representation learning using convolutional layers and TPP to convert inputs into fixed-dimensional vectors, allowing it to classify different renditions of the same composition as well as different music [18].*

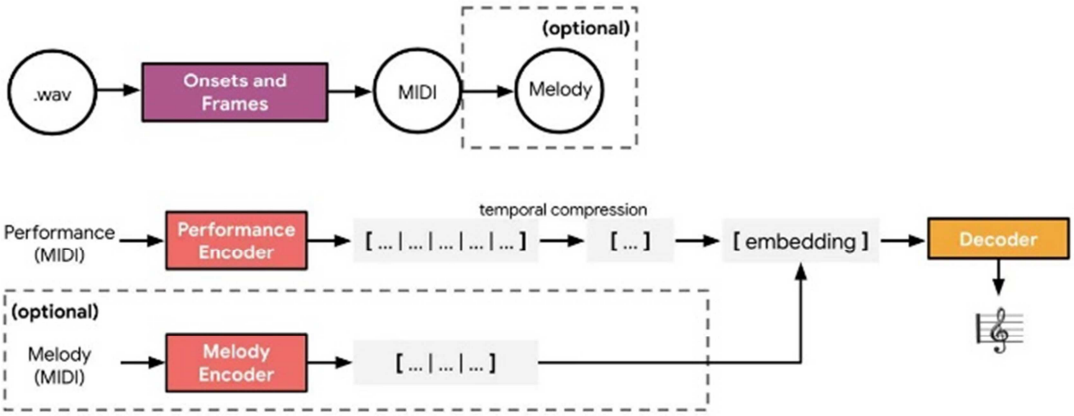

Figure 3. *The wav data files are initially transcribed into midi format using the onsets and frames frame-work. These are later encoded into performance representations for input. The aggregated out-put of the performance encoder, along with a melody embedding if desired, forms a comprehensive representation of the entire performance, which is utilized for inference by the transformer decoder [19].*

4. Models

The models used to compute similarity are mainly based on feature extraction algorithms such as CPCP, HPCP, MFCC, and then utilize similarity measurement methods such as SNF to calculate the similarity between music pieces [17]. This approach is based on matrix calculation, where the similarity features of music can be quantified as numerical values. Such methods are applied in tasks like cover song recognition and music classification. The development of artificial intelligence technology, based on large datasets, allows for the extraction of local or temporal features from data using neural networks. Others proposed a CNN-based method to calculate similarity. This method treats different renditions of the same song as one class and different songs as different classes, constructing a dataset of audio recordings [18]. Low-level descriptors are extracted using the CQT method, and a convolutional neural network is trained to learn this feature, outputting Temporal Pyramid Pooling (TPP) as shown in Figure 2, transforming music into fixed-dimensional features. Unlike traditional methods based on similarity matrices, neural network-based methods generally require large datasets, and different network models can be used to learn music similarity features for different purposes. For example, CNN is good at handling local features, RNN and LSTM can handle temporal features, transformer can extract global features of style as shown in Figure 3 [19]. By extracting more useful features, neural networks bring more possibilities to similarity calculation. In

addition to the similarity calculation methods based on music feature and deep learning, there are also other approach-es and models for calculating music similarity. These include estimating music similarity based on music context data [10], calculating music similarity based on music tag information, computing music similarity through user behavior analysis (by analyzing user preferences), and computing music similarity based on bag-of-words model and TF-IDF (converting music into text information and performing calculation).

5. Application

Similarity calculations are widely used in music for various purposes such as music recommendation, music classification and tagging automation, cover song and hummed song recognition, artist identification, music sampling, music generation, copyright protection, and infringement detection. Here are a few commonly used directions briefly introduced.

5.1. Music Recommendation

As everyone more or less enjoys music today, music recommendations have become common in our daily lives. For example, when one opens certain music apps, based on our searches, likes, favorites, shares, and other actions, the app will start recommending music that one may like. Not only that, it will also recommend the most listened to and trending music to us. All of this is achieved by analyzing and calculating similar music based on our behavior as users. One of the most common and widely used methods in the early days was collaborative filtering (CF). This method predicts user preferences based on the preferences of similar users as well as the user's own past preferences [20]. As shown in Figure 4, the left side of the figure, music recommendations are based on user ratings, while on the right side, recommendations are based on a set of music that is related. If a user likes one of the songs in the set, the whole set of music will be recommended to the user. In addition, in recent years, con-tent-based filtering (CBF) methods have also gained popularity [21].

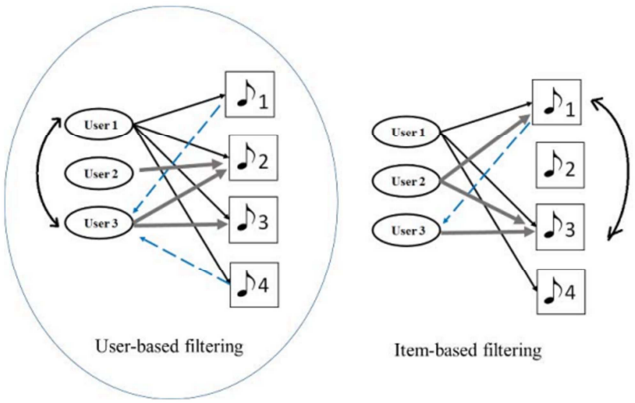

Figure 4. *Two methods of collaborative filtering [21].*

5.2. Artist Identification

Previously, a trained LDA generative model was used for clustering and similarity function to calculate the similarity between music [22]. It was applied to artist identification, proposing a clustering of artists from different topics using this model and similarity function. Due to the difficulty of processing a large database, clustering algorithms were proposed to identify singer similarity and group them, thus helping to handle large datasets [23].

5.3. Music Classification and Automated Tagging

Traditional machine learning classifiers, machine learning methods, and deep learning methods were used for automatic classification of music genres [24]. The advantages and disadvantages of different methods were also analyzed and compared. Additionally, the study found that users tend to browse music based on a common genre rather than the similarity of artists, further confirming the meaningfulness of the proposed automatic music classification method. Of course, with the development of artificial intelligence, there have been numerous techniques for classifying and labelling music genres, which are not listed here.

5.4. Music Generation

Music generation has become increasingly popular in recent years, and Google has conducted research on music generation, believing that music generation systems will become increasingly important in our lives. Music can be generated using artificial intelligence technologies such as machine learning, but it first requires a music corpus or specific music styles for the machine learning model to learn from in order to generate the desired music. This also indicates that regardless of the type of music generation, a relevant concept, i.e., similarity [25]. Additionally, it is considered a core indicator of successful music generation, as mentioned in [25]. If the similarity is too high, there is a suspicion of plagiarism. Therefore, finding a balance between similarity and innovation is an important challenge.

5.5. Humming Recognition

Many music apps now have the function of recognizing songs by humming, which is also related to computing similarity. When researchers do Humming recognition, factors such as the original music rhythm, tempo, and duration may vary. These all need to be normalized using algorithms to find matching music segments. The commonly used techniques for measuring the similarity of music segments are DTW and its variations, such as edit distance on real sequence (EDR) and edit distance with real penalty (ERP), which measure the distance between them [23].

5.6. Others

In this era of information explosion, computing similarity has not only been widely applied in music, but also plays an important role in other fields. For example, computing

similarity is used to measure text similarity (search engines, text classification, machine translation, article plagiarism), image similarity, video similarity, audio similarity, voiceprint similarity, etc. It is also used in commercial fields, such as online shopping apps which calculate similarity based on user's browsing history to recommend similar products. Similarly, some short video apps calculate similarity history, preferences (likes, favorites), and areas of interest to recommend videos of similar types to users. Additionally, some question-and-answer communities and social media platforms calculate similarity based on user's browsing and search records, social relationships, and content of interest, and then recommend it to users.

6. Limitations & Future Outlooks

With the development of artificial intelligence technology, the calculation of music similarity is no longer limited to traditional mathematical formula calculations based on manually extracting audio signal features. Instead, there is now an additional meth-od of using machine learning classifiers or algorithms and models to calculate music similarity using manually extracted features as input data. Whether it is traditional mathematical calculation or the use of novel artificial intelligence technology, both methods have their advantages and limitations. This article will summarize some limitations of current music similarity calculation and application, aiming to better assist researchers who are interested in understanding the shortcomings of current technology and are seeking to break through and innovate.

First, let's discuss the traditional and most common distance calculation methods: cosine distance (comparing the similarity of two vectors by calculating the cosine of the angle between them after feature extraction) and Euclidean distance (calculating the distance between two points in space). If one wants to calculate the similarity of audio feature sequences, cosine distance is more suitable because calculating Euclidean distance for high-dimensional data can be computationally expensive and time-consuming. However, using cosine distance to measure similarity also has limitations. Firstly, it requires a significant amount of time to extract music features before calculating the cosine distance. Secondly, when using cosine similarity, only the direction of the vectors is considered, neglecting the vector's length, which leads to inaccuracies in reflecting the degree of music similarity. Thirdly, cosine similarity only considers the issue of music feature vectors, ignoring semantic information of the music, resulting in less accurate calculations. Lastly, cosine distance focuses solely on overall similarity of the music, neglecting the correlation between the music's internal and contextual elements. In conclusion, relying solely on cosine distance for judgement is definitely not accurate enough. Therefore, it is necessary to incorporate more music features and algorithms to improve its accuracy, which could involve advanced artificial intelligence techniques. Secondly, it is also possible to use SVM, a commonly used machine learning algorithm model for classification, to compute music semantic indicators. It has significant advantages in solving high-dimensional pattern recognition problems and when there are fewer samples available. However, it should be noted that when training on large-scale music datasets, support vector machine requires high computational and storage requirements, leading to increased runtime and space complexity. Therefore, it may be worth considering other machine learning methods or hybrid models to calculate music similarity. Deep learning models can also be used and trained on existing music label data. However, regardless of the model chosen, it needs to be trained on a large amount of data to achieve higher accuracy. Training a model can be challenging and requires a computer with high configuration. In addition to acquiring a large amount of data, it is also necessary to determine if the data can be used and whether there are any copyright infringement concerns.

In conclusion, no single method for calculating similarity is perfect. Some methods are suitable for calculating similarity in high-level music features, while others are suitable for calculating similarity in mid-level music features. Therefore, combining different methods to calculate music similarity can be a good choice, resulting in potentially more accurate similarities. For the future, the author hope to see the following advancements in future research on computational music similarity: Firstly, the author hopes to have a dedicated dataset in the future that contains high-quality music of various styles, with numerous similar music segments. Secondly, the author hopes to witness the creation of new, more versatile, and accurate models by combining different algorithms. Thirdly, the author hopes for the development of a mature computational music similarity system that can determine whether music is plagiarized, quantify the extent of plagiarism, and have commercial potential for legal purposes.

7. Conclusion

To sum up, this article summarizes some commonly used principles and processes for extracting music similarity, as well as models and approaches for calculating music similarity, and the application and limitations of calculating music similarity. Through research, it is found that there is currently no mature system that can determine the degree of similarity between two pieces of music without consuming manpower. The author believes that one major reason is that the available music dataset is not of high quality and diversity, and the existing models are too complex to be widely applicable. The author hopes to have a high-quality music dataset in the future, with diverse styles and numerous similar music examples. Additionally, it is hoped for models with broader applicability and higher calculation accuracy. The vision is to see a mature detection system in the future that can be commercialized and provide assistance in legal matters, and so on. The hope is that this article can provide some references for scholars who want to understand music similarity and those who want to overcome the limitations of current music similarity calculations.

References

[1] Funk, T., A Musical Suite Composed by an Electronic Brain: Reexamining the Illiac Suite and the Legacy of Lejaren A. Hiller Jr., *Leonardo Music Journal*, *28*, 2018, pp. 19-24.

[2] Meng Tongtong, Development and Application Analysis of Computer Music, *Music Space and Time*, (8x), 2018, pp. 74.

[3] Cheng, C., Comparative Research on Song Similarity Based on Deep Learning, Doctoral Dissertation, Beijing: Beijing University of Posts and Telecommunications, 2020.

[4] Li, W., Li, Z., & Gao, Y. Understanding digital music: A review of music information retrieval techniques, *Fudan Journal (Natural Science Edition)*, *57(3)*, 2018, pp. 271-313.

[5] Chen, N., Li, W., & Xiao, H. Fusing similarity functions for cover song identification, *Multimedia Tools and Applications, 77,* 2018, pp. 2629-2652.

[6] Seyerlehner, K., Widmer, G., & Pohle, T. Fusing block-level features for music similarity estimation. In Proc. of the 13th Int. Conference on Digital Audio Effects (DAFx-10), 2010, pp. 225-232.

[7] Flexer, A., & Grill, T. The problem of limited inter-rater agreement in modelling music similarity, *Journal of new music research, 45(3),* 2016, pp. 239-251.

[8] Knees, P., Schedl, M., Knees, P., & Schedl, M. (2016). Introduction to music similarity and retrieval. *Music Similarity and Retrieval*, 1-30.

[9] Berenzweig, A., Logan, B., Ellis, D. P., & Whitman, B. A large-scale evaluation of acoustic and subjective music-similarity measures, *Computer Music Journal,* 2004, pp. 63-76.

[10] Knees, P., & Schedl, M. A survey of music similarity and recommendation from music context data, *ACM Transactions on Multimedia Computing, Communications, and Applications (TOMM), 10(1)*, 2013, pp. 1-21.

[11] Chen, N., Li, M., & Xiao, H. Two-layer similarity fusion model for cover song identification, *EURASIP Journal on Audio, Speech, and Music Processing, 2017 (1)*, 2017, 1-15.

[12] Seyerlehner, K., Schedl, M., Pohle, T., & Knees, P. Using block-level features for genre classification, tag classification and music similarity estimation, *Submission to Audio Music Similarity and Retrieval Task of MIREX, 2010 (2)*, 2010, pp. 3.

[13] Fujishima, T. Realtime Chord Recognition of Musical Sound: a System Using Common Lisp Music. In Proc. of the International Computer Music Conference 1999, 1999, pp. 464-467.

[14] Serra, J., Serra, X., & Andrzejak, R. G. Cross recurrence quantification for cover song identification. *New Journal of Physics, 11(9)*, 2009, 093017.

[15] Chen, N., Downie, J. S., Xiao, H. D., & Zhu, Y. Cochlear pitch class profile for cover song identification, *Applied Acoustics, 99,* 2015, pp. 92-96.

[16] Gómez, E. Tonal description of music audio signals. PhD Thesis, Universität Pompeu Fabra, Barcelona, 2006.

[17] Chen, N., & Xiao, H. D. Similarity fusion scheme for cover song identification. *Electronics Letters, 52(13),* 2016, pp. 1173-1175.

[18] Yu, Z., Xu, X., Chen, X., & Yang, D. Temporal Pyramid Pooling Convolutional Neural Network for Cover Song Identification. In Proc. of the International Joint Conference on Artificial Intelligence 2019, 2019, pp. 4846-4852.

[19] Choi, K., Hawthorne, C., Simon, I., Dinculescu, M., & Engel, J. Encoding musical style with transformer autoencoders. In Proc. of the International Conference on Machine Learning, 2020, pp. 1899-1908.

[20] Shakirova, E. Collaborative filtering for music recommender system. In Proc. of the 2017 IEEE Conference of Russian Young Researchers in Electrical and Electronic Engineering (EIConRus), 2017, pp. 548-550.

[21] Schedl, M. Deep learning in music recommendation systems. *Frontiers in Applied Mathematics and Statistics*, 2019, pp. 44.

[22] Knees, P., & Schedl, M. Music similarity and retrieval: an introduction to audio-and web-based strategies, Vol. 36. Heidelberg: Springer, 2016.

[23] Murthy, Y. S., & Koolagudi, S. G. Content-based music information retrieval (cb-mir) and its applications toward the music industry: A review. *ACM Computing Surveys (CSUR), 51(3)*, 2018, pp. 1-46.

[24] Ndou, N., Ajoodha, R., & Jadhav, A. Music genre classification: A review of deep-learning and traditional machine-learning approaches. In 2021 IEEE International IOT, Electronics and Mechatronics Conference (IEMTRONICS), 2021, pp. 1-6.

[25] Herremans, D., Chuan, C. H., & Chew, E. A functional taxonomy of music generation systems. *ACM Computing Surveys (CSUR), 50(5)*, 2017, pp. 1-30.