Proceedings of the 2023 International Conference on Software Engineering and Machine Learning
DOI: 10.54254/2755-2721/8/20230220

Comparison of multiple machine learning algorithms for music genre classification

Diwen Deng[1, †], Yiwu Gu[2, †], Yiyi Zhu[3, 4, †]

[1] School of Electronic and Electrical Engineering, Shanghai University of Engineering Science, Shanghai, 201600, China
[2] School of Mathematical Science, East China Normal University, Shanghai, 200241, China.
[3]College of Computer Science and Technology/College of Artificial Intelligence /College of Software, Nanjing University of Aeronautics and Astronautics, Liyang, Jiangsu, 213300, China

[4] zhuyiyi @nuaa.edu.cn
[†]These authors contributed equally

Abstract. With the fast advance of the Internet and the continuous improvement of computer technology, speech recognition has been applied in many fields, and speech recognition has broad prospects for development. Music and audio classification technology can add category labels to music based on music content, which is of great significance in the research and application of efficient organization, retrieval and recommendation of music resources. In order to efficiently classify audio from massive online music data and help users to obtain the most suitable music style, a deep learning classification algorithm based on convolutional neural network (CNN) is proposed. To examine its effectiveness, it is compared with traditional machine learning algorithm. First, the original music data set was preprocessed and then feature extraction was carried out to obtain music features and transform them into spectral maps. Traditional machine learning model and deep learning component model were used for simulation experiments. The testing accuracy of the deep learning model is up to 92%, verifying the model's superiority.

Keywords: Music genre classification, machine learning, classification

1. Introduction

With the advent of the digital age, multimedia information technology has progressively penetrated into every aspect of people's life. Image, audio, and video gradually become the main forms of multimedia digital information. Among them, the digital music resources show an explosive growth. Massive music resources stimulate music users to produce different music retrievals. Therefore, how to quickly find the favorite music types from the massive digital music resources is the research focus in the field of music retrieval. During retrieval, people classify different music by distinguishing the unique musical characteristics of different music. Music genre is an important label to describe the characteristics of music. No matter what genre, the tone, harmony, frequency and other characteristics of music are extremely varied, while different users have different love degree for various genres of music. Therefore, people choose to classify music from the perspective of musical genre.

DOI: 10.54254/2755-2721/8/20230220

Early musical classifications were made through professionals' listening and annotating [1]. Later music classification was done by manually extracting features and then using machine learning methods [2]. However, these methods have great defects. First, if people just manually extract the features or classify them, it is difficult to guarantee the validity and accuracy of the result. Second, traditional machine learning methods are unable to process large-scale data and perform poorly on classification problems. But nowadays, the update speed of music resources far exceeds the classification speed of traditional music classification methods. Therefore, the current popular method is the deep learning methods. Deep learning methods can enhance the performance of music genre classification. However, the computational consumption of it is high, and the model design is also complex. So, it needs to be further combined with music information to improve the overall performance of classification.

Audio classification tasks need to be applied to multiple disciplines, such as voice signal processing, probability, artificial intelligence, statistics, etc. In recent years, the research on audio classification and identification has made long-term development and new breakthroughs, but the focus of the research has always been around the extraction of more effective audio features and the exploration of more appropriate classification model.

Since the 1990s, people began to use music classifiers for classification, such as the famous Music Genome Project, where a group of musicians and loving technicians got together to divide each piece of music into many factors, called "genes", and used these factors to recommend favorite music to users. Matityaho B et al use fast Fourier and log transformation to extract music frequency features, and use multi-layer neural network as a decision system to classify classical and pop music [3]. Jiang DN et al proposed using spectral divergence as a musical feature and a relative spectral distribution rather than the average spectral envelope, which is more suitable for extracting musical features [4]. The researchers found that the spectrum map of music data can effectively depict musical characteristics, and the spectrum map can well depict the frequency and time of music. Later, people began to use the ideas of neural networks to study musical classification. Costa Y et al. used the spectrum map and found that the local binary mode (LBP) is better applied to music classification [5]. Sarkar R et al. used empirical pattern decomposition (EMD) to extract the features of music information and classify it using a multi-layer perceptron network [6]. Cahyani D E et al. used the principal component analysis method (PCA), using K proximity, Bayesian network and sequence minimum optimization algorithm [7].

2. Method

2.1. Dataset

The famous GITZAN datasets are used in the case study. There are ten genres, including disco, hip-hop, pop, classical, country, rock, blues, jazz, reggae, and metal, each kind has one hundred sound clips each are included in the data collection, which has 1,000 recordings averaging 30 seconds. Before training the classification model, the raw data in the audio sample is converted into a more meaningful representation, and the audio footage is transformed from.au format to.wav so that it can be read by the python waveform function.

2.2. Mel-frequency cepstral coefficients

In Figure1, the spectrum is drawn over time to depict different types of musical frequencies. The most used in music classification research is the Mel-frequency cepstral coefficients (MFCC). For human, there are different sensitivity to different sound waves, where 200~5000 Hz is the most sensitive frequencies. However, the perception of sound waves of other frequencies is weaker, so people use mathematical models to simulate this unique physiological structure. The spectrum shows that the characteristics of music of different genres vary with time changing, which is used as the basis for classification.

DOI: 10.54254/2755-2721/8/20230220

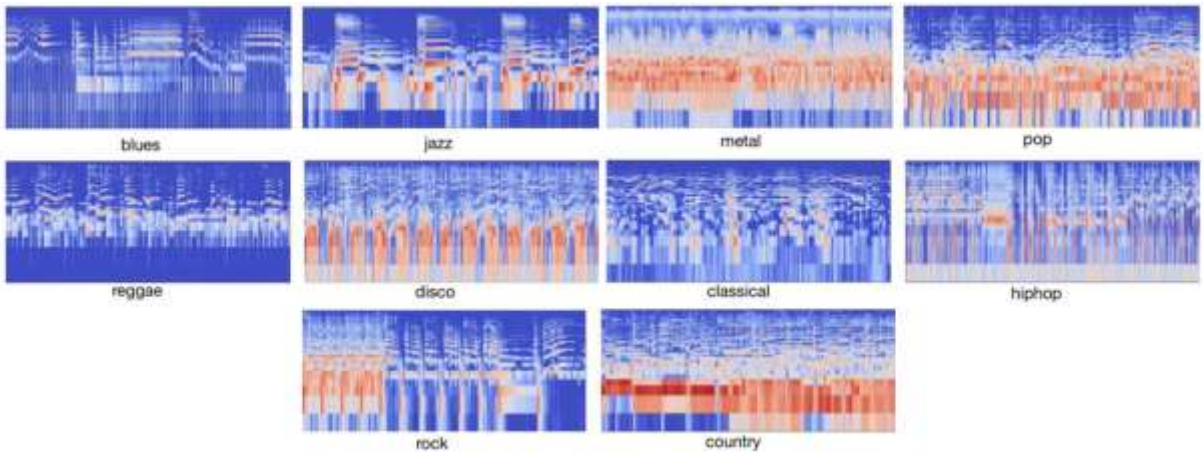

Figure 1. Frequency demonstration of different music genres.

2.3. Machine learning models

2.3.1 Naïve Bayes. It is an effective algorithm, based on Bayes theorem and independent assumption of characteristic conditions [8]. Bayes theorem is shown in (1):

$$P(A_i|B) = \frac{P(B|A_i)P(A_i)}{\sum_j P(B|A_j)P(A_j)} \tag{1}$$

where, $P(A_i)$ denotes prior probability and $P(B|A_i)$ denotes conditional probability. Naive Bayes model estimates conditional and prior probability under the premise of independent characteristic conditions. And then the model according to the conditional and prior probability to calculate the posterior probability of new samples. In reality, however, the characteristic conditions are not always independent. The false premise thus leads to misclassification.

2.3.2 K Nearest Neighbors. KNN uses supervised learning to classify new instances based on the nearby training examples that are already present in the latent space. The principle of it is to take the distance of the data point and categories of its K neighbors. Prediction points are classified into the class with the minimum distance.

The major aspects influencing the classification effect of KNN model are the value of K and the calculation method of distance. The result will affected by the noise if K is too small, and the approximation inaccuracy of the model will rise if K is too large [9]. In addition, the distance measure directly determines the classification results. The commonly used distance measures are Euclidean metric, Manhattan distance and Mahalanobis distance [10].

Formulas of distance measurement is as follows:

$$d_1 = \sqrt{\sum_{i=1}^{n}(x_i - y_i)^2} \tag{2}$$

$$d_2 = |x_i - x_j| + |y_i - y_j| \tag{3}$$

$$d_3 = \sqrt{(X-\mu)^T S^{-1}(X-\mu)} \tag{4}$$

where, d_1 represents Euclidean Metric, d_2 represents Manhattan distance and d_3 represents Mahalanobis distance.

The value of k in this experiment is 3, and Euclidean Metric is chosen as the distance formula.

DOI: 10.54254/2755-2721/8/20230220

2.3.3 Decision tree. The model determines the likelihood that the predicted value of net present value is greater than or equal to zero, based on the known occurrence probabilities of various events.

This tree-like structure's basic morphology is shown in:

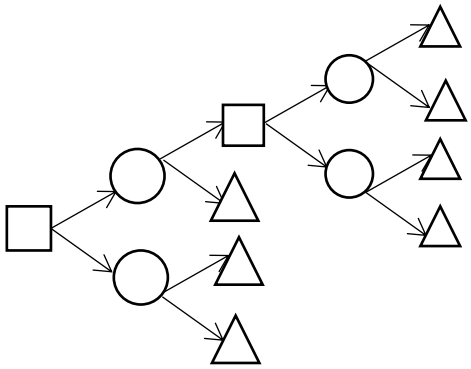

Figure 2. Structural demonstration of decision tree.

Here, a circle denotes a chance node, a square represents a decision node, and a triangle is an end node (leaf node).

Each classification starts from the root node to test the feature of the predicted points. According to the test result, the predicted point is assigned to its child nodes, where each child node corresponds to a value of the feature. Such a recursive testing instance and assign, until it reaches the leaf nodes. Finally, the predicted points are classified into the class of the leaf nodes.

2.3.4 Random forest. It is to train multiple decision tree models, and use the classification results of multiple trees to integrate the judgment to give the final result. This model reflects the idea of ensemble learning [10]. Experiments have shown that the depth of the forest affects the classification results. In this experiment, the depth value is 20.

2.3.5 Support vector machine. SVM is a binary classifier grounded on supervised learning, which can be estimated by constructing kernel function. The basic principle of SVM is to establish an optimal decision-making hyperplane $w^T x + b = 0$ to separate samples of two categories, thereby maximizing the distance between samples closest to the hyperplane and thereby improving the generalization ability of a classification model [10].

In order to achieve the multi-class classification, the multi-class classification strategy used in this experiment is OVR (One Vs Rest).

2.3.6 Logistic regression. The core idea of logistic regression is to fit a logical function to predict the probability of an event, sigmoid function is generally used as the prediction function [11].

2.3.7 Deep learning models. The neural network data set is composed of abundant synapse configurations [12]. With reference to the MLP-like architecture, Fully Connected Neural Network can be divided into input layer, hidden layer with the ReLU activation function, and output layer with the Softmax. Neurons of each layer of the fully connected neural network are connected to each other. Its basic morphology is shown in Figure 3 a). Here, 4 network layers are constructed in this experiment. Moreover, to avoid over-fitting, each layer needs dropout. The rendering of the dropout is shown in Figure 3 b).

DOI: 10.54254/2755-2721/8/20230220

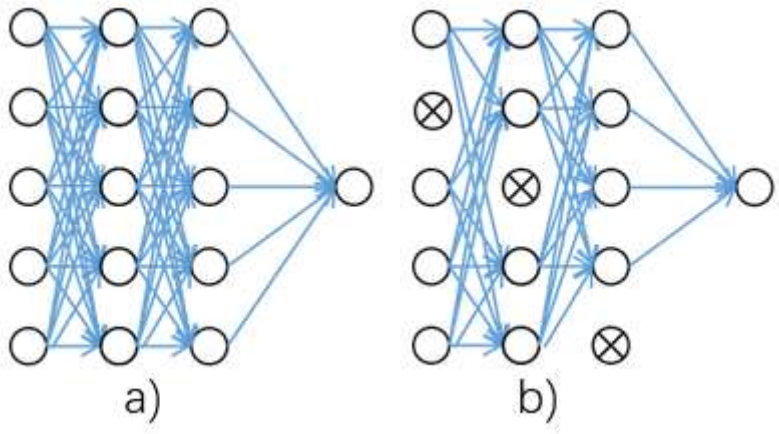

Figure 3. Demonstration of fully connected layer and dropout.

In this experiment, each layer uses 20% dropout. To train the model, the experiment chooses the Adam optimizer and sets the epoch to 60.Moreover, the sparse_categorical_crossentropy function is leveraged to estimate the loss.

3. Result

Comparing the accuracy of each model is presented in Table 1. The training progress is demonstrated in Figure 4. The result demonstrates that the classification accuracy of these 7 classification models is above 50% in this data set. Among these, FCNN's classification was the best, with the accuracy of 91.6% in this data set. Secondly are KNN, Random Forest and SVM. The accuracy of them respectively is 90.5%, 87.4% and 86.0%

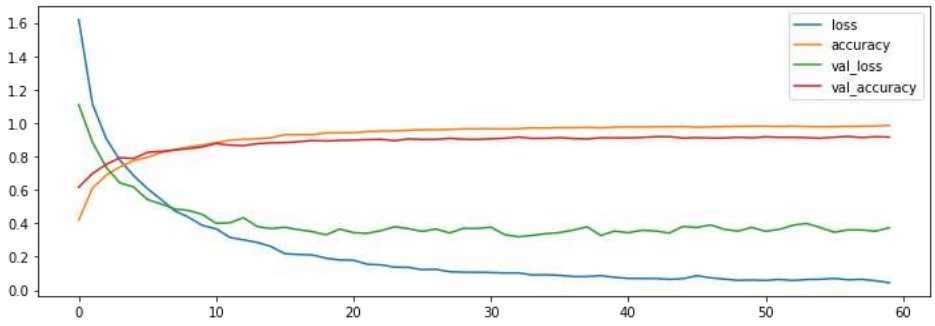

Figure 4. Loss curve and accuracy during training.

Table 1. Result of different models.

Classification Models	Accuracy	Precision	Recall	F1
Naive Bayes	51.4%	52.1%	51.6%	49.9%
KNN	90.5%	90.6%	90.4%	90.5%
Decision Trees	63.0%	63.7%	63.1%	63.3%
Random Forest	87.4%	87.4%	87.4%	87.3%
SVM	84.8%	84.9%	84.9%	84.8%
Logistic Regression	72.6%	72.7%	72.8%	72.7%
FCNN	91.6%			

DOI: 10.54254/2755-2721/8/20230220

4. Discussion

This experiment trained the classification model for music genre recognition leveraging six conventional algorithms plus a fundamental deep learning technique. The outcomes of the experiment demonstrate that the deep learning model performs more accurately in this classification issue. There are some plausible explanations based on prior research. Firstly, the deep learning model's network topology exhibits high non-linearity, enabling the fitting of complex functions as well as the ability to thoroughly abstract the features without overfitting. Secondly, the multi-level structure created by the deep learning model learns more valuable features during the layer-by-step feature transformation process, improving the accuracy of the classification results [13].

Despite the fact that the deep learning model in this experiment has a better classification effect compared to the traditional machine learning model, it still has some limitations. Compared with other deep learning algorithms, the number of Fully Connected Neural Network layers is low in this experiment, and each layer adopts a fully-connected layer. To ensure that, while completing the feature extraction, the tested target is not submerged in any other unnecessary background. Using 10-fold cross validation is likely to achieve a better classification effect (93.12%) [14].

At present, the classification of music genres is more accurate, but there is room for improvement. Future work can further improve the deep learning algorithm to improve the performance. The classification results can create better commercial value by more accurately analyzing user preferences and predicting trends in music popularity [15].

5. Conclusion

In this paper, Naive Bayes, K Nearest Neighbors, DecisionTrees, Random Forest, Support Vector Machines, Logistic Regression, Fully Connected Neural Network and deep learning algorithm are used to conduct music type classification experiment of audio data model. By comparing experimental results, the deep learning model reaches the highest accuracy of 92%. In general, the deep learning model achieves a good classification of audio types. Audio classification technology based on machine learning has very important research significance in daily life, and has broad application prospects in the future. This paper can also try to learn the original audio signal of music directly, and try to consider the original musical audio signal as the gateway to the network to extract the music features, in such a way as to avoid the loss of musical data during the conversion of the sound spectrum.

References

[1] Vishnupriya, S., & Meenakshi, K. (2018). Automatic music genre classification using convolution neural network. In 2018 international conference on computer communication and informatics (ICCCI) , 1-4.

[2] Xu, Y., & Zhou, W. (2020). A deep music genres classification model based on CNN with Squeeze & Excitation Block. In 2020 Asia-Pacific Signal and Information Processing Association Annual Summit and Conference (APSIPA ASC), 332-338.

[3] Matityaho, B., & Furst, M. (1995). Neural network based model for classification of music type. In Eighteenth Convention of Electrical and Electronics Engineers in Israel, 3-4.

[4] Jiang, D., Lu, L., Zhang, H., Tao, J., & Cai, L. (2002). Music type classification by spectral contrast feature. In Proceedings. IEEE International Conference on Multimedia and Expo, 1, 113-116.

[5] Costa, Y. M., Oliveira, L. S., Koerich, A. L., Gouyon, F., & Martins, J. G. (2012). Music genre classification using LBP textural features. Signal Processing, 92(11), 2723-2737.

[6] Sarkar, R., & Saha, S. K. (2015). Music genre classification using EMD and pitch based feature. In 2015 Eighth International Conference on Advances in Pattern Recognition (ICAPR), 1-6.

[7] Cahyani, D., & Nuzry, K. (2019). Trending topic classification for single-label using multinomial naive bayes (MNB) and multi-label using k-nearest neighbors (KNN). In 2019 4th International Conference on Information Technology, Information Systems and Electrical Engineering (ICITISEE), 547-552.

DOI: 10.54254/2755-2721/8/20230220

[8] Liu, Z., Zhang, Q. M., Lü, L., & Zhou, T. (2011). Link prediction in complex networks: A local naïve Bayes model. Europhysics Letters, 96(4), 48007.
[9] Gang, Z., Shi-kui, P., Hui, R., et, al. (2010). A general introduction to estimation and retrieval of forest volume with remote sensing based on KNN. Remote sensing technology and application, 25(3), 430-437.
[10] Chillara, S., Kavitha, A. S., Neginhal, S. A., Haldia, S., & Vidyullatha, K. S. (2019). Music genre classification using machine learning algorithms: a comparison. Int Res J Eng Technol, 6(5), 851-858.
[11] Cokluk, O. (2010). Logistic Regression: Concept and Application. Educational Sciences: Theory and Practice, 10(3), 1397-1407.
[12] Scabini, L. F., & Bruno, O. M. (2023). Structure and performance of fully connected neural networks: Emerging complex network properties. Physica A: Statistical Mechanics and its Applications, 128585.
[13] LeCun, Y., Bengio, Y., & Hinton, G. (2015). Deep learning. nature, 521(7553), 436-444.
[14] Li, L. (2021). Learning Recommendation Algorithm Based on Improved BP Neural Network in Music Marketing Strategy. Computational Intelligence and Neuroscience, 1-10.
[15] Pandeya, Y. R., & Lee, J. (2021). Deep learning-based late fusion of multimodal information for emotion classification of music video. Multimedia Tools and Applications, 80, 2887-2905.