Classification and Popularity Assessment of English Songs Based on Audio Features

Tao Wang
School of Computer Science and Technology
Shandong University of Technology
Zibo, China

Abstract:- Nowadays a large amount of new music emerges every year. How to properly categorize music for quick browsing and retrieval by users and evaluate music popularity based on audio features is an important research topic. In this study, the decision tree model is used to classify music styles on a dataset consisting of audio features of 4802 songs from 2008-2017. Then, the number of music listening in the dataset was used as an indicator to assess the popularity of songs. By comparing the training results of different Machine Learning algorithms on the dataset, Gradient Boosting Regressor is chosen to be used in this case, and the relative importance of different audio features on the popularity of songs was calculated with this model.

Keywords:- *Audio Features, Machine Learning, Classification.*

I. INTRODUCTION

With the rapid development and popularity of the Internet and information technology, online music has become an essential form of entertainment in people's daily lives. Music websites and applications based on streaming technology have also become the primary channels for people to access music. At the same time, a large amount of new music emerges every year. It becomes an important research topic to classify these music appropriately for users to browse and retrieve them quickly and evaluate whether users will welcome the songs based on the features in the music.

Music is a more complex audio information than speech, containing various elements such as human voice, musical instruments, nature sounds, and noise. In the early days of music information processing research, the focus was on music recognition and retrieval methods. It was not until the 1990s, with the rise of Internet technology, that the field of music classification algorithms came to the forefront. Matisyahu et al. proposed a method in 1995 to preprocess audio information using the Fourier transform and then classify it using artificial neural networks [1]. In 1996 Wold et al. proposed to use the mean, variance, and autocorrelation correlation coefficients as features to classify audio signals using the KNN algorithm [2]. In 2002, Tzanetakis et al. used timbre pitch and rhythm as features to classify music with 61% classification accuracy [3]. In 2012, the Google Brain project used a single amount of computing resources to train a deep neural network (DNN), which achieved a significant breakthrough in speech recognition and image processing. In addition to the classical algorithms listed above, there are many practical music classification algorithms. These methods are based on extracting features that reflect the essential properties of music, designing high-performance classifiers, and optimizing the classification results.

Meanwhile, with the significant increase in the number of online music releases each year, how to predict the popularity of music and push music on this basis has become an important area that affects the activity of music website users [4-5]. Auditing every music piece without a purpose will undoubtedly add a lot of unnecessary time costs for music users. Since many users browse and enjoy music works on electronic music platforms every day, the resulting massive amount of recorded data on users' browsing collections and listening to music is an essential guide to music trends and users' preferences [6].

This study is based on 17.7K English song data from 2008-2017 and the track metrics compiled by The Echo Nest on the Kaggle platform. The dataset provides several audio features, including acousticness, danceability, energy, instrumentalness, liveness, tempo, and valence, as well as label data such as music classification and the number of music listening. This study uses the above data set to train the music classifier with the Decision Tree algorithm. At the same time, the number of music listening is the most famous indicator of music works, and the evaluator of user preferences is constructed using music feature data.

II. MUSIC STYLE CLASSIFICATION

In this study, the decision tree model is used to classify music styles on a dataset consisting of audio features of 4802 songs from 2008-2017. The Decision Tree model used for the classification problem generalizes the classification rules from the training dataset and is a supervised learning method. The decision tree model has a tree-like structure and represents classifying data based on features. The advantage is that the model is readable, and the classification is faster than other commonly used algorithms. Although there is an infinite number of conditional probability models based on the class division in the feature space, during the training process of the decision tree, the model that fits the training data well and has excellent predictive power for the unknown data should be selected.

ISSN No: 2456-2156

Before starting to train the classification model, all the song dataset is divided into the training data set and the test data set, in which the training data set accounts for 70% of the total. The statistical analysis of the data labels of different music styles in the training dataset shows far more data classified as rock music than hip-hop music. The unbalanced classification problems will potentially skew the model's ability to distinguish between classes. Therefore, the ratio of two types of music data in the training dataset is adjusted to match roughly.

The typical construction methods of decision trees are mainly ID3, C4.5, and CART. Here, the CART algorithm is used to construct the classification model, and the Gini index is used as the criterion for feature selection. When training the model with default hyperparameters, the model is evaluated using the test data set, and the results obtained are shown in Table 1.

TABLE 1- Classification results with default hyperparameters

	Precision	Recall	F1-score
Hip-Hop	0.77	0.78	0.78
Rock	0.78	0.77	0.77
macro avg	0.77	0.77	0.77
weighted avg	0.77	0.77	0.77

Precision, recall and F1 score are used to measure the accuracy of the classification model. Hip-Hop and Rock is the category label for music styles. The reported averages include macro average (averaging the unweighted mean per label), weighted average (averaging the support-weighted mean per label).

The performance of the classification model significantly depends on the value of hyperparameters. To further improve the performance of the classification model, GridSearchCV technique was used to tune the model parameters, which is the process of performing hyperparameter tuning to determine the optimal values for the given model.

Predefined hyperparameters values are passed to the algorithm. It tries all the combinations of the values passed in the dictionary and evaluates the model for each combination using the Cross-Validation method. After we get accuracy/loss for every combination of hyperparameters, choose the one with the best performance to evaluate the classification model. After hyperparameter tuning, model parameters such as max_depth, min_impurity_decrease, and min_samples_leaf are set to 10, 0.005, and 10, correspondingly. The results obtained from the model evaluation are shown in Table 2.

TABLE 1- Classification results with default hyperparameters

	Precision	Recall	F1-score
Hip-Hop	0.82	0.79	0.80
Rock	0.80	0.83	0.81
macro avg	0.81	0.81	0.81
weighted avg	0.81	0.81	0.81

By comparing Table 1 and Table 2, it can be seen that the classification performance of the model has been improved. Meanwhile, the graphical representation of the decision tree model obtained from the training is shown in Fig.1

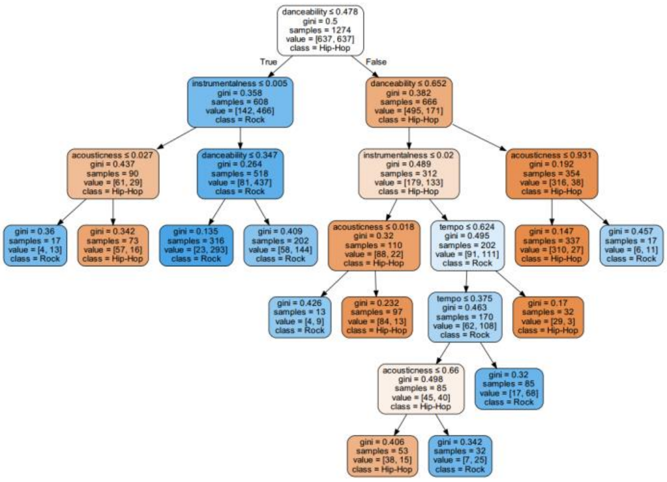

Fig.1 Graphical representation of the decision tree model

III. SONG POPULARITY ASSESSMENT

In this study, the features such as acousticness, danceability, energy, instrumentalness, liveness, tempo, valence and bit rate in the audio dataset are used to predict the number of future listens of a song, i.e., the song's popularity. The audio datasets are split into training and test sets using a random permutation cross-validator.

In order to build a more accurate assessment model, eight regression analysis algorithms are used here to construct a regression model of song popularity. The input and output data of the models were normalized. The hyperparameters of each regression algorithm model were empirically given initial custom values. The max_iter of MLP Regressor is limited to 2000. The Gradient Boosting Regressor, Ada Boost Regressor, and Bagging Regressor model are limited to have 1000 n_estimators. The prediction results of different models were evaluated with the training and test sets. RMSE and max_error are used as the metrics for evaluating the prediction models, respectively.

It can be seen that the Train_RMSE and Test_RMSE of BaggingRegressor and Gradient Boosting Regressor are relatively close and have better results than the other models, as shown in Fig.2

However, the training time of Gradient Boosting Regressor is significantly lower than that of Bagging Regressor. Therefore, Gradient Boosting Regressor is chosen as the prediction model in this study.

ISSN No: 2456-2156

	Name	Parameters	Train_RMSE	Test_RMSE	Test_max_error	Comsumed Time
3	MLPRegressor	{'activation': 'tanh', 'alpha': 0.0001, 'batch...	0.077495	0.075569	-0.296309	0.197916
8	SVR	{'C': 1.0, 'cache_size': 200, 'coef0': 0.0, 'd...	0.041146	0.040534	-0.264695	0.007902
4	AdaBoostRegressor	{'base_estimator': None, 'learning_rate': 0.01...	0.006198	0.014862	-0.389472	5.11973
5	BaggingRegressor	{'base_estimator': None, 'bootstrap': True, 'b...	0.007534	0.01457	-0.284976	18.897847
2	GradientBoostingRegressor	{'alpha': 0.9, 'ccp_alpha': 0.0, 'criterion': ...	0.007816	0.011937	-0.313575	5.806884
0	LinearRegression	{'copy_X': True, 'fit_intercept': True, 'n_job...	0.018344	0.010262	-0.330055	0.000984
6	LinearSVR	{'C': 1.0, 'dual': True, 'epsilon': 0.0, 'fit_...	0.018535	0.010221	-0.330111	0.086758
7	NuSVR	{'C': 1.0, 'cache_size': 200, 'coef0': 0.0, 'd...	0.018149	0.01011	-0.329401	6.87198
1	DecisionTreeRegressor	{'ccp_alpha': 0.0, 'criterion': 'mse', 'max_de...	0.018474	0.010087	-0.3305	0.002333

Fig.2 Comparation of different machine learning algorithms

GridSearchCV technique was used to search for better model hyperparameters for Gradient Boosting Regressor, to improve the model accuracy further. The model parameters selected through a limited number of searches are:" learning_rate=0.01, max_depth=15, min_samples_leaf=10, min_samples_split=10". Training deviance during the model training process is shown in Fig.3

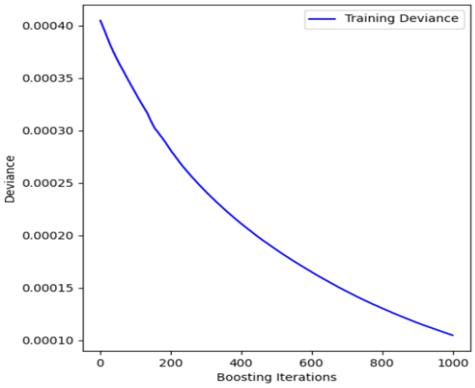

Fig.3 Training deviance during the model training process

Meanwhile, the relative importance of different music features in the song dataset was evaluated by the feature importance attribute of Gradient Boosting Regressor algorithm, as shown in Fig.4

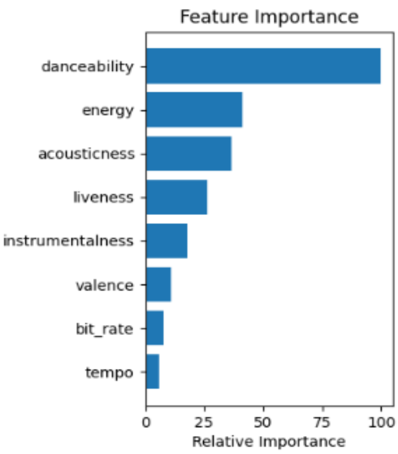

Fig.4 Relative importance of different music features

IV. SUMMARY

In this study, 17.7K English song data on the Kaggle platform were analyzed in-depth by machine learning algorithms. First, a classifier of song styles was constructed by a decision tree algorithm and several audio features in the dataset.

Then, the number of songs listens to in the dataset was used as an indicator to assess the popularity of songs. By comparing the training results of eight standard regression analysis algorithms on this dataset, the Gradient Boosting Regressor was selected as the predictor in this case, and the relative importance of different audio features on the popularity of songs was calculated with this model.

REFERENCES

[1]. Matityaho B , Furst M . Neural network based model for classification of music type[C]// Electrical & Electronics Engineers in Israel, Eighteenth Convention of. IEEE, 1995.
[2]. Wold E, Blum T, Keislar D, et al. Content-based classification, search, and retrieval of audio[J]. IEEE multimedia, 1996, 3(3): 27-36.
[3]. Tzanetakis G , Cook P . Musical genre classification of audio signals[J]. IEEE Transactions on Speech & Audio Processing, 2002, 10(5):293-302.
[4]. C Hamm. Putting Popular Music in its Place: Some thoughts on the measurement of popularity in music. 1995.
[5]. Monechi B , Gravino P , Servedio V , et al. Significance and popularity in music production[J]. Royal Society Open Science, 2017, 4(7):170433.
[6]. Martin-Gutierrez D , Penaloza G H , Belmonte-Hernandez A , et al.
[7]. A Multimodal End-To-End Deep Learning Architecture for Music Popularity Prediction[J]. IEEE Access, 2020, PP(99):1-1.