**２０２１　年度　　 卒業論文**

卒業論文題目

|  |
| --- |
| ラウンドロビン方式の負荷分散による  楽曲分類システムの設計と開発 |

　指導教員　　　　　　鷹野　孝典　教授

神奈川工科大学　情報工学科

　　　　　学籍番号　　　　1821144

　　　　　学生氏名　　　　　 吉井 智哉

　提出日　2021年　12月23日　指導教員　　印

　受理日　2021年　１2月23日　情報工学科長　　印

論文要旨

目次

[1. まえがき 1](#_Toc90481872)

[**1.1** はじめに 1](#_Toc90481873)

[**1.2** 課題と提案内容 1](#_Toc90481874)

[**1.3** 本論文の構成 1](#_Toc90481875)

[**2.** 関連研究 2](#_Toc90481876)

[**2.1** 動画の印象評価データセット構築とその特性の調査 2](#_Toc90481877)

[**2.2** 2](#_Toc90481878)

[**2.3** ニューラルネットワークを用いた音楽の自動ジャンル分類 2](#_Toc90481879)

[**3.** 提案システム 3](#_Toc90481880)

[**3.1** 提案システムの概要 3](#_Toc90481881)

[**3.2** 提案システムの実行手順 3](#_Toc90481882)

[**4.** 実装システム 4](#_Toc90481883)

[**4.1** 実装システムの概要 4](#_Toc90481884)

[**4.2** 実装システムの処理手順 4](#_Toc90481885)

[**4.3** 環境構築 4](#_Toc90481886)

[**4.4** 実装システムの構築 5](#_Toc90481887)

[**5.** 実験 8](#_Toc90481888)

[6. むすび 9](#_Toc90481893)

# まえがき

## はじめに

　インターネットの発展や，スマートフォンのようなコンピュータの普及によりSNS(Social Networking Service)が大きく発展した，これに伴って音楽投稿型のSNSも同様に発展を遂げた．そのため誰でも音楽をSNSなどに投稿できるようになっており，SNSなどに投稿される楽曲が莫大な量になってきている．

　また， 機械学習でよく利用されるCNN(Convolutional Neural Network)だがこの処理を行うことでかかる負荷が高いとされている．

　本研究では楽曲コンテンツにCNNを用いてジャンル推定を行うため，楽曲SNSに投稿される大量楽曲コンテンツのジャンル推定処理に高負荷がかかると予測されるが，この処理を複数のサーバで分散処理することで処理速度の向上を実現する．

## 課題と提案内容

　SNSに投稿される大量のデータをジャンル推定処理する場合，1台のみで処理を行うと処理速度に限界が生じる．この処理速度をできるだけ早くすることがサービス向上につながると考えられる．

本研究では，複数のサーバーにこの処理を分散することで処理速度の向上を目的としたシステムを提案する．

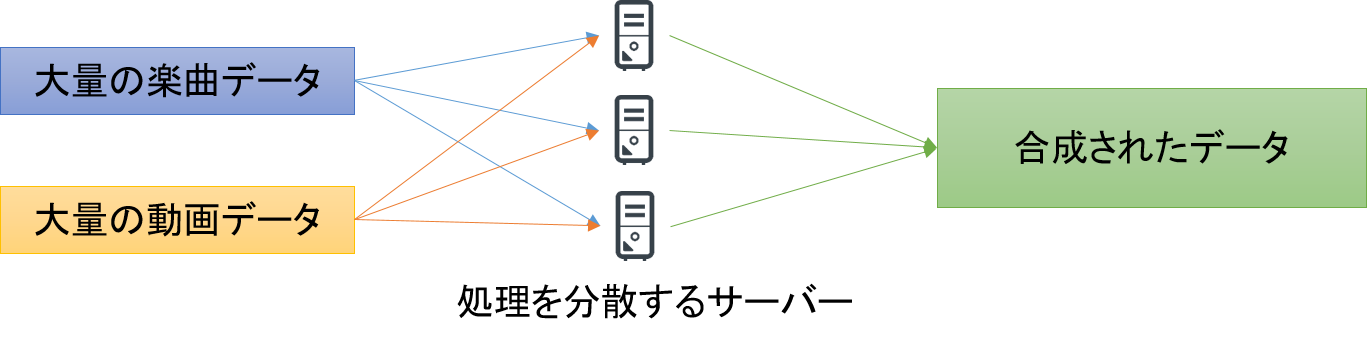


図 1.1

## 本論文の構成

本論文では全6章で構成されている．本章では研究の背景や動機，目的について述べる第2章では関連研究について述べる．第3章では本研究の提案方式とその特徴である本システムの内容，第4章では提案方式を構築した本システムの設計及び実験システムの実装について述べる．第5章では本システムの実験第6章では謝辞および今後の展望を述べる

# 関連研究

## 動画の印象評価データセット構築とその特性の調査

音楽に対する印象評価に関する研究は多数なされている中で大野直樹，中村聡史，山本岳洋，後藤真孝らはそうした研究を促進するためのデータセットも様々なものが構築されている一方で音楽と映像が同期して提示される音楽動画を対象とした印象評価に関する研究は、ほとんどなされていないという．この研究では500曲の音楽動画のサビ区間を対象とし，音楽のみ，動画のみ，音楽と映像の見合わせという3つのタイプの評価対象コンテンツを用意する．また，このコンテンツに対して8軸の印象評価を行ってもらうことで，メディアの及ぼす影響を明らかにする．さらに，これまで大野らが行ってきた，音楽動画全体に対する印象評価と，この研究で収集した音楽動画のさび区間に対する印象評価とを比較することで印象評価において注意すべき点についての考察を行った．

## 

## ニューラルネットワークを用いた音楽の自動ジャンル分類

　スタンダードMIDIファイル形式の楽曲に対してその部分情報からその楽曲のジャンル推定をし，特徴をフィードバックするジャンル学習支援システムを構築することを最終目的とし．現在までに，4ジャンル計120曲をニューラルネットワークに学習させ，その学習させた局のジャンルを同定させることができている．さらに，評価実験として被験者にジャンル推定を行わせ，学習させたニューラルネットワークとの，推定率の比較を行い，本システムのほうが高い推定率を得たとしている．続いて学習させたニューラルネットワークに対して中間層の各ノードと連結しているリンクの重みを見比べることにより，専門家による意味解釈を行った．その結果，5つの中間層に対して，因子を抽出できた，これにより教育利用のためのジャンル学習支援システムの構築の可能性が示されたとしている．

# 提案システム

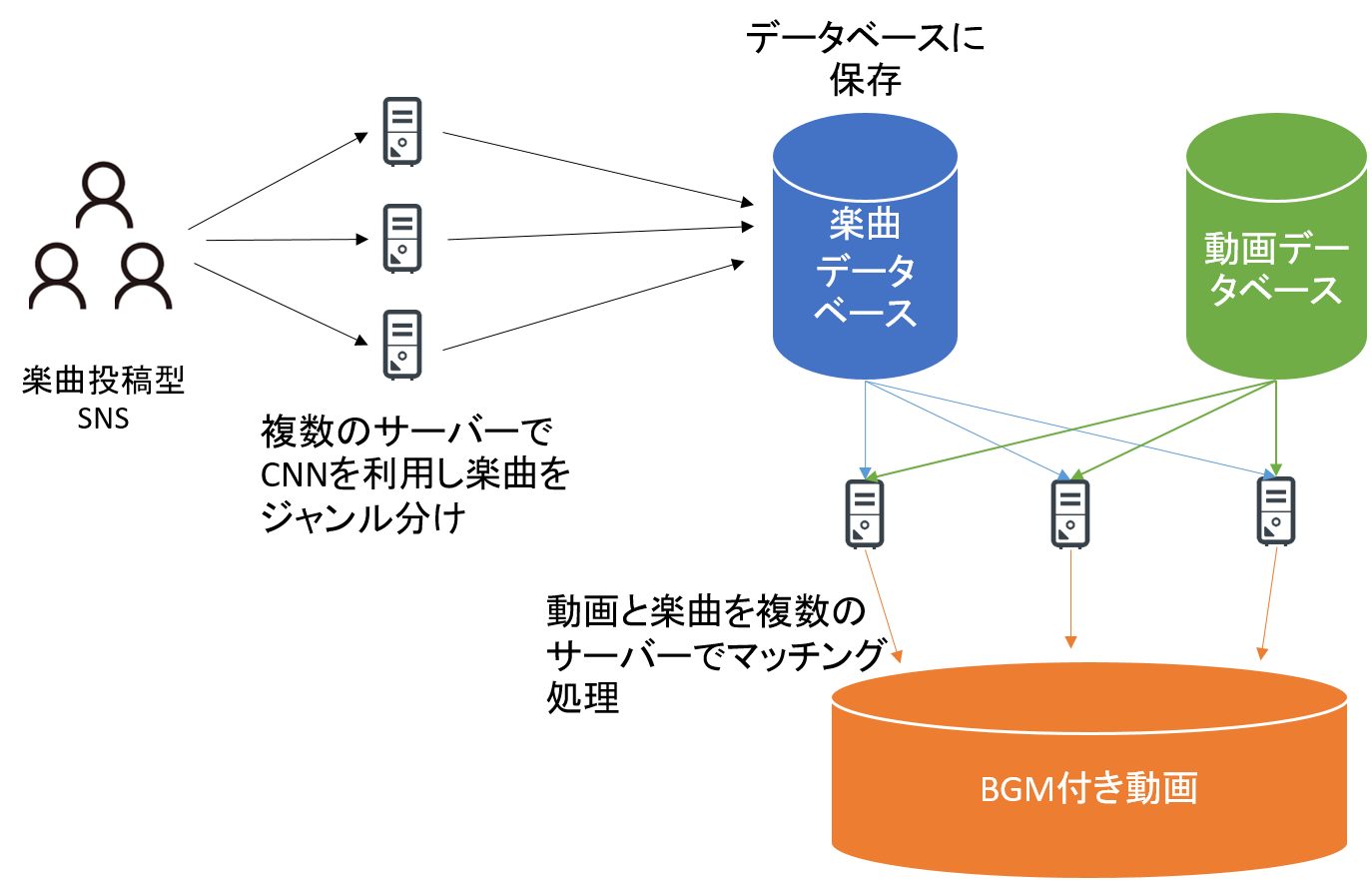
## 提案システムの概要

本研究では，SNSなどに投稿される大量の楽曲を複数のサーバーで並列にCNN(Convolutional Neural Network)を利用し楽曲のジャンル分類を行い，その楽曲と動画をマッチングするシステムを提案する．

本提案システムではジャンル分類に機械学習を利用するため，システムの処理に時間がかかる．加えてSNSなどに投稿される大量のデータを処理するため，なおさら時間がかかることが予測される．このような処理を1つのサーバーで処理を行おうとすると処理時間が長くなってしまう．そこで処理するサーバーを複数にし，並列処理でシステムの処理を行うことで単純計算でも2台で半分になる．これによりシステム全体の処理時間を減らすことが可能である．またジャンル分類を行った楽曲コンテンツと動画コンテンツのマッチングも並列で行うことで更なる処理時間の短縮を実現する．

## 提案システムの実行手順

1. 分散処理するためのサーバを数台用意する
2. 用意したサーバでstep2~step6を処理
3. 楽曲SNSに投稿された楽曲を保存
4. 保存した楽曲の周波数を分析
5. 分析した周波数をスペクトログラムにして可視化
6. スペクトログラムを機械学習し楽曲ジャンルを推定



# 実装システム

## 実装システムの概要

実装システムでは楽曲のジャンル推定処理を3台のRaspberry piに分散処理することで処理の負荷分散を実現している．

実装システムで分散処理に使用したサーバ3台，負荷分散機1台はそれぞれRaspberry Pi 4B 4GBを使用する．また本システムのメインプログラムはプログラミング言語Pythonで記述している．

## 実装システムの処理手順

　実装システムの処理手順は下記に示す．

1. 楽曲の周波数を分析する
2. 分析した周波数を画像として可視化する
3. FMA\_SMALLデータセットを用いてジャンル推定モデルの作成
4. ジャンル推定モデルとstep2の画像を用いてジャンル推定
5. step1からstep4の処理をRaspberry Pi3台で分散処理

## 環境構築

提案システムのプロトタイプを実装した開発環境を表4.1に示す．提案システムを作成するために用いたパッケージを表4.2に示す．

表 4.1システム開発環境

|  |  |  |
| --- | --- | --- |
| 名称 | バージョン | 利用用途 |
| Windows10 | 20H2 | OS |
| Raspberry Pi | 4B 4GB | 分散処理に使用 |
| Raspberry Pi OS |  | Raspberry Pi OSのOS |
| Python | 3.7.3 | 開発言語 |
| FFmpeg | 4.4 | 楽曲を再生・変換する |

表 4.2　システム構築に用いたパッケージ

|  |  |  |
| --- | --- | --- |
| ライブラリ名 | バージョン | 備考 |
| pydub | 0.25.1 | 音楽ファイルの読み込みに使用 |
| pandas | 1.3.4 | データ解析を容易にする |
| numpy | 1.21.3 | 機械学習の計算を早く効率的に行えるようにする |
| matplotlib | 3.4.3 | Pythonでグラフを描画する |
| seaborn | 0.11.2 | matplotlibの内部で動く．少ないコードで洗礼された図を描く |
| llvmlite | 0.32.1 | 後述するnumbaとlibrosaのインストールに必要 |
| numba | 0.49.0 | Pythonの関数を高速にするライブラリ |
| LibROSA | 0.8.1 | 音楽と音声の解析のためのPythonパッケージ．スペクトル解析，テンポの分析，画像出力など音楽の分析に必要な機能があらかじめ実装されている． |
| TensorFlow | 2.1.1 | Googleが開発した機械学習のソフトウェアライブラリ |
| keras | 2.3.1 | Pythonでかかれた  TensorFlow上で実行可能な高水準のニューラルネットワークライブラリ |

実装システムで利用したPythonパッケージのインストール方法を表4.3に示す

表 4.3　Pythonパッケージのインストール

|  |
| --- |
| $ sudo apt-get update  $ sudo apt-get upgrade  $pip3 install pandas == 1.3.4  $pip3 install numpy == 1.21.3  $pip3 install matplotlib == 3.4.3  $pip3 install seaborn == 0.11.2  $ sudo apt install llvm-7  $LLVM\_CONFIG=/usr/bin/llvm-config-7  $pip3 install llvmlite==0.31.0  $pip3 install numba==0.48.0  $pip3 install colorama==0.3.9  $pip3 install librosa==0.6.3  $pip3 install keras == 2.3.1  $pip3 install tensorflow == 2.3.1  $pip3 install h5py == 2.10.0 |

## 実装システムの構築

　実装システムの構築手順を4.4.1からで説明する

### numpyを利用した楽曲分析

　機械学習を行うために楽曲の周波数を分析しスペクトログラムとして可視化する必要がある．スペクトログラムとは図4.1のような縦軸に周波数を，横軸に時間をとり色によってどの程度の周波数が多く出ているか表したものである．図4.1のスペクトログラムで示された楽曲では100Hz～900Hzの音が多く使われていることがわかる．

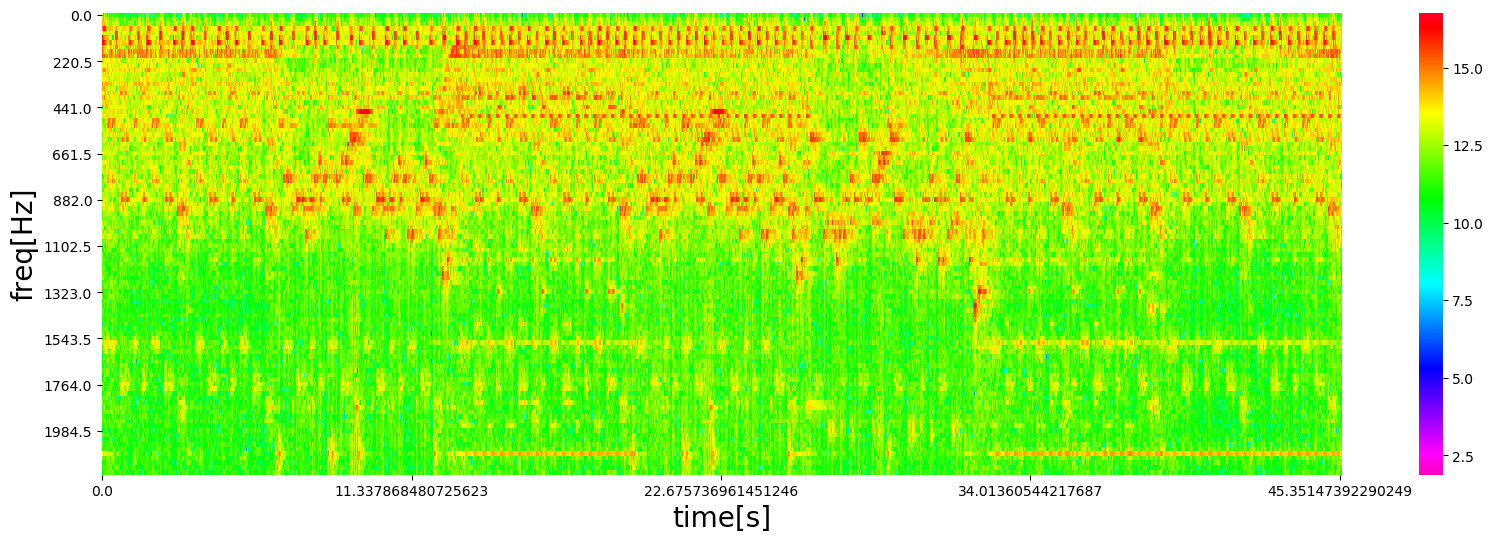


図 4.1スペクトログラム

本項では楽曲をスペクトログラムとして出力するプログラムheatmap.pyについて説明する．

まずは使用すPythonパッケージのインポートをする．

表 4.4パッケージのインポート

|  |
| --- |
| from pydub import AudioSegment  import os  import numpy as np  import pandas as pd  import matplotlib.pyplot as plt  import seaborn as sns |

楽曲を分析するには楽曲を読み込む必要がある楽曲を読み込むプログラムは表4.4に示す．

表 4.5楽曲の読み込み

|  |
| --- |
| sound = AudioSegment.from\_file(‘filename’) |

この時filenameは解析したい楽曲のファイルの名前を入れる．

読み込んだ楽曲をnumpyで読み込みステレオ音声で片側のみの音声を抽出する。

表 4.6　numpyでステレオ音声の抽出

|  |
| --- |
| samples = np.array(sound.get\_array\_of\_samples())  sample = samples[::sound.channels] |

表 4.7フーリエ変換

|  |
| --- |
| spec = np.fft.fft(sample) |

表 4.8スペクトログラムのラベル表示

|  |
| --- |
| w = 2000　#窓幅  s = 500　#刻み |

表 4.9スペクトログラムの計算

|  |
| --- |
| #スペクトル格納用  ampList = []  #刻みずつずらしながら窓幅分のデータをフーリエ変換する  for i in range(int((sample.shape[0]- w) / s)):  data = sample[i\*s:i\*s+w]  spec = np.fft.fft(data)  spec = spec[:int(spec.shape[0]/2)]  spec[0] = spec[0] / 2  ampList.append(np.abs(spec)) |

これで楽曲に対する処理を終えたので，matplotlibとseabornを利用して実際にスペクトログラムとして可視化する．

表 4.10スペクトログラムとして可視化

|  |
| --- |
| df\_amp = pd.DataFrame(data=ampList, index=time, columns=freq)  a = plt.figure(figsize=(20, 6))  l = sns.heatmap(data=np.log(df\_amp.iloc[:, :100].T),  xticklabels=1000,  yticklabels=10,  cmap=plt.cm.gist\_rainbow\_r,  )  l.set\_ylabel("freq[Hz]",fontsize = 20)  l.set\_xlabel("time[s]",fontsize = 20)  a.savefig("%s\/%s.png"%(path, name),bbox\_inches="tight") |

これでスペクトログラムとして楽曲を可視化することができる．

### kerasを利用してジャンル推定モデルの作成と楽曲ジャンル推定

本項ではPythonパッケージkerasを用いてジャンル推定モデルの作成を行う．

ジャンル推定モデルの作成ではFMA\_SMALLデータセットを用いる．またジャンル推定におけるジャンルの種類はElectric，Folk，HipHop，International，Latin，Metal，Noise，Pop，Rock，Punkの10ジャンルをとする．

始めに使用したPythonパッケージを表4.11に示す

表 4.11使用したパッケージ

|  |
| --- |
| import librosa  import pandas as pd  import numpy as np  import matplotlib.pyplot as plt  import os  from PIL import Image  import pathlib  import csv  import time  # Preprocessing  from sklearn.model\_selection import train\_test\_split  from sklearn.preprocessing import LabelEncoder, StandardScaler  #Keras  import keras  from keras.models import Sequential,load\_model  from keras import models  from keras import layers  import warnings  warnings.filterwarnings('ignore') |

モデル作成に使用するデータセットをそれぞれのジャンルフォルダに格納した後，格納した楽曲から特徴を抽出しスペクトログラムに変換し，スペクトログラムにしたものをimg\_dateフォルダの各ジャンルフォルダに保存

表 4.12

|  |
| --- |
| cmap = plt.get\_cmap('inferno')  plt.figure(figsize=(10,10))  genres = 'Electric Folk HipHop International Latin Metal Noise Pop Rock Punk'.split()  for g in genres:  pathlib.Path(f'img\_data/{g}').mkdir(parents=True, exist\_ok=True)  for filename in os.listdir(f'{g}'):  songname = f'{g}/{filename}'  y, sr = librosa.load(songname, mono=True, duration=5)  plt.specgram(y, NFFT=2048, Fs=2, Fc=0, noverlap=128, cmap=cmap, sides='default', mode='default', scale='dB');  plt.axis('off');  plt.savefig(f'img\_data/{g}/{filename[:-3].replace(".", "")}.png')  plt.clf() |

スペクトログラムにしたものからメル周波数ケプストラム係数(MFCC)，Spectral Centroid，，ZeroCrossingRate，Spectral Roll-offといった特徴を抽出し，抽出したものをcsvファイルとして保存．

表 4.13特徴を抽出したものをcsvに保存

|  |
| --- |
| header = 'filename chroma\_stft spectral\_centroid spectral\_bandwidth rolloff zero\_crossing\_rate'  for i in range(1, 21):  header += f' mfcc{i}'  header += ' label'  header = header.split()  file = open('data.csv', 'w', newline='')  with file:  writer = csv.writer(file)  writer.writerow(header)  genres = 'Electric Folk HipHop International Latin Metal Noise Pop Rock Punk'.split()  for g in genres:  for filename in os.listdir(f'{g}'):  songname = f'{g}/{filename}'  y, sr = librosa.load(songname, mono=True, duration=30)  chroma\_stft = librosa.feature.chroma\_stft(y=y, sr=sr)  spec\_cent = librosa.feature.spectral\_centroid(y=y, sr=sr)  spec\_bw = librosa.feature.spectral\_bandwidth(y=y, sr=sr)  rolloff = librosa.feature.spectral\_rolloff(y=y, sr=sr)  zcr = librosa.feature.zero\_crossing\_rate(y)  mfcc = librosa.feature.mfcc(y=y, sr=sr)  to\_append = f'{filename} {np.mean(chroma\_stft)} {np.mean(spec\_cent)} {np.mean(spec\_bw)} {np.mean(rolloff)} {np.mean(zcr)}'  for e in mfcc:  to\_append += f' {np.mean(e)}'  to\_append += f' {g}'  file = open('data.csv', 'a', newline='')  with file:  writer = csv.writer(file)  writer.writerow(to\_append.split()) |

表 4.14Pandasでデータを分析

|  |
| --- |
| data **=** pd**.**read\_csv('data.csv')  data**.**head()　#上から5件を表示 |

表 4.15不必要なコラムを削除

|  |
| --- |
| data **=** data**.**drop(['filename'],axis**=**1) |

表 4.16ラベルをエンコード

|  |
| --- |
| genre\_list **=** data**.**iloc[:, **-**1]  encoder **=** LabelEncoder()  y **=** encoder**.**fit\_transform(genre\_list)  scaler **=** StandardScaler()  X **=** scaler**.**fit\_transform(np**.**array(data**.**iloc[:, :**-**1], dtype **=** float) |

表 4.17トレーニングデータセットとテストデータセットに分ける

|  |
| --- |
| X\_train, X\_test, y\_train, y\_test = train\_test\_split(X, y, test\_size=0.2) |

表 4.18kerasを利用したジャンル分類モデルの構築

|  |
| --- |
| model **=** models**.**Sequential()  model**.**add(layers**.**Dense(256, activation**=**'relu', input\_shape**=**(X\_train**.**shape[1],)))  model**.**add(layers**.**Dense(128, activation**=**'relu'))  model**.**add(layers**.**Dense(64, activation**=**'relu'))  model**.**add(layers**.**Dense(10, activation**=**'softmax'))  model**.**compile(optimizer**=**'adam',  loss**=**'sparse\_categorical\_crossentropy',  metrics**=**['accuracy']  history **=** model**.**fit(X\_train,  y\_train,  epochs**=**20,  batch\_size**=**128)  test\_loss, test\_acc **=** model**.**evaluate(X\_test,y\_test)  print('test\_acc: ',test\_acc) |

以上でモデルの作成ができた．ここからジャンル推定のプログラムを下記に示す．

|  |
| --- |
| predictions = model.predict(X\_test)  predictions[0].shape  np.sum(predictions[0])  prediction = np.argmax(predictions[0])  print(genres[prediction]) |

これでジャンル推定することができた．

# 実験



# むすび