

Redes Neurais em predição de gênero musical

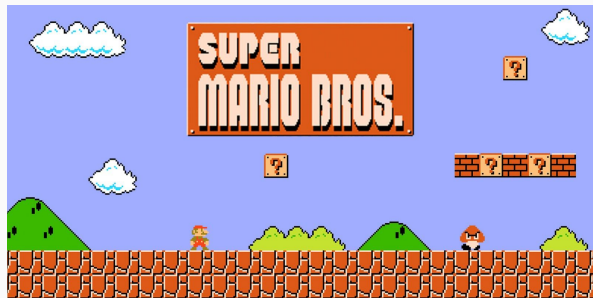
Tiago Costa Carvalho - 11315101

Gustavo Fernandes Carneiro de Castro - 11369684

Introdução

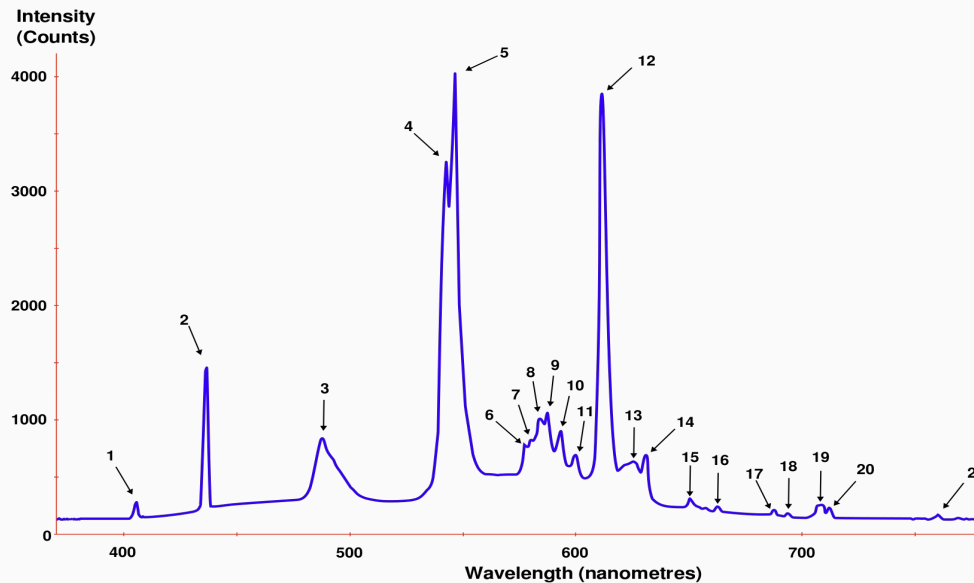
Objetivo

Utilizar-se de um perceptron multicamadas (MLP) para prever o gênero musical de um conjunto de músicas selecionadas arbitrariamente.



Introdução

Para isto, foram obtidos coeficientes MFC (*mel-frequency cepstrum*), através da biblioteca *librosa*, do *python*. Além disso, também foi usado *scikit learn* e o *pandas* para as análises e utilização dos algoritmos.



Metodologia

Gêneros escolhidos

foram selecionados arbitrariamente 10 gêneros musicais para serem estudados:

- Clássico (C),
- Eletrônica (E)
- Heavy Metal (HM)
- Jazz (J)
- Pop (P)
- Rock (R)
- Rhythm & Blues/Soul (R&B/ReB)
- Rap e derivados (RAP)
- Retro Gaming (RG)
- Samba (S)

Para cada gênero, foi selecionado, em média 22 músicas, totalizando 205 músicas no conjunto de dados final.

librosa

Através desta biblioteca, foram retirados os coeficientes usados neste estudo ao inserir os arquivos *.mp3* das músicas escolhidas arbitrariamente para representar o gênero musical.

Utilizando o método *librosa.feature.mfcc()*, extrai-se os *Mel-frequency cepstral coefficients* (MFCCs) de um dado *.mp3*. Esses coeficientes representam; de forma simples, o espectro de força de uma onda sonora ao longo do tempo.

librosa

Com isso, para cada coeficiente i , $i = \{1, \dots, 20\}$, foi calculado o valor deste em todos os instantes t de cada música. Depois, foi calculada a média do coeficiente i em todos os instantes t , finalizando com um total de $n = 20$ atributos para cada música.

Instantes t

Coefficients

[[-568.85394 -568.85394 -568.85394 ... -568.85394 -568.85394 -568.85394]

[0. 0. 0. ... 0. 0. 0.]

[0. 0. 0. ... 0. 0. 0.]

...

[0. 0. 0. ... 0. 0. 0.]

[0. 0. 0. ... 0. 0. 0.]

[0. 0. 0. ... 0. 0. 0.]

(20, 7414)

Análise

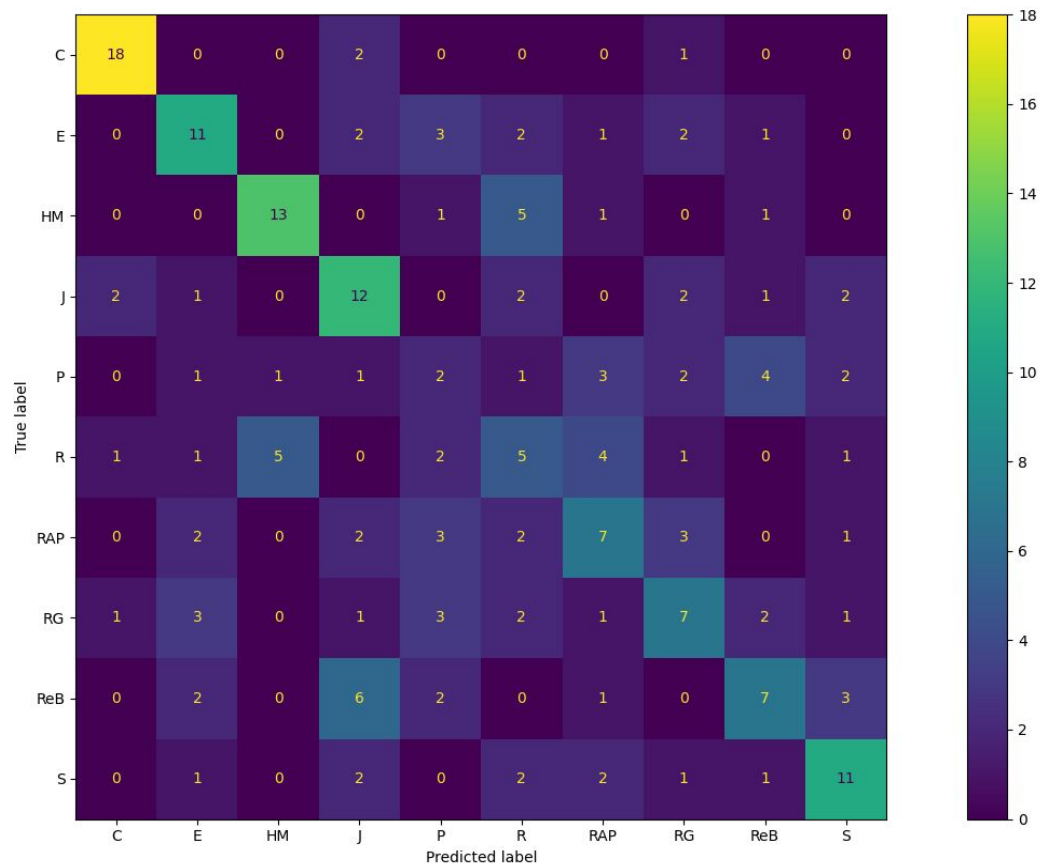
Foi utilizado o método *GridSearchCV()* para automatizar os testes para diversas opções de parâmetros. Entre eles:

- Função de ativação: *'identity'* - função linear, *'logistic'* - função logística sigmoidal, *'tanh'* - função da tangente hiperbólica e *'relu'* - ativação linear retificada
- Quantidade de camadas ocultas, e quantidade de neurônios em cada uma (por ex: [200,200])
- taxa de aprendizado inicial: 0.001, 0.01, ...
- algoritmo de solução para otimização de pesos: *lbfgs*, *sgd*, *adam*
- número máximo de iterações: 1000, 10000,
- aleatorização da ordem dos dados de entrada: *sim*, *não*.

Além disso, foi utilizado um cross-validation com 5 folds.

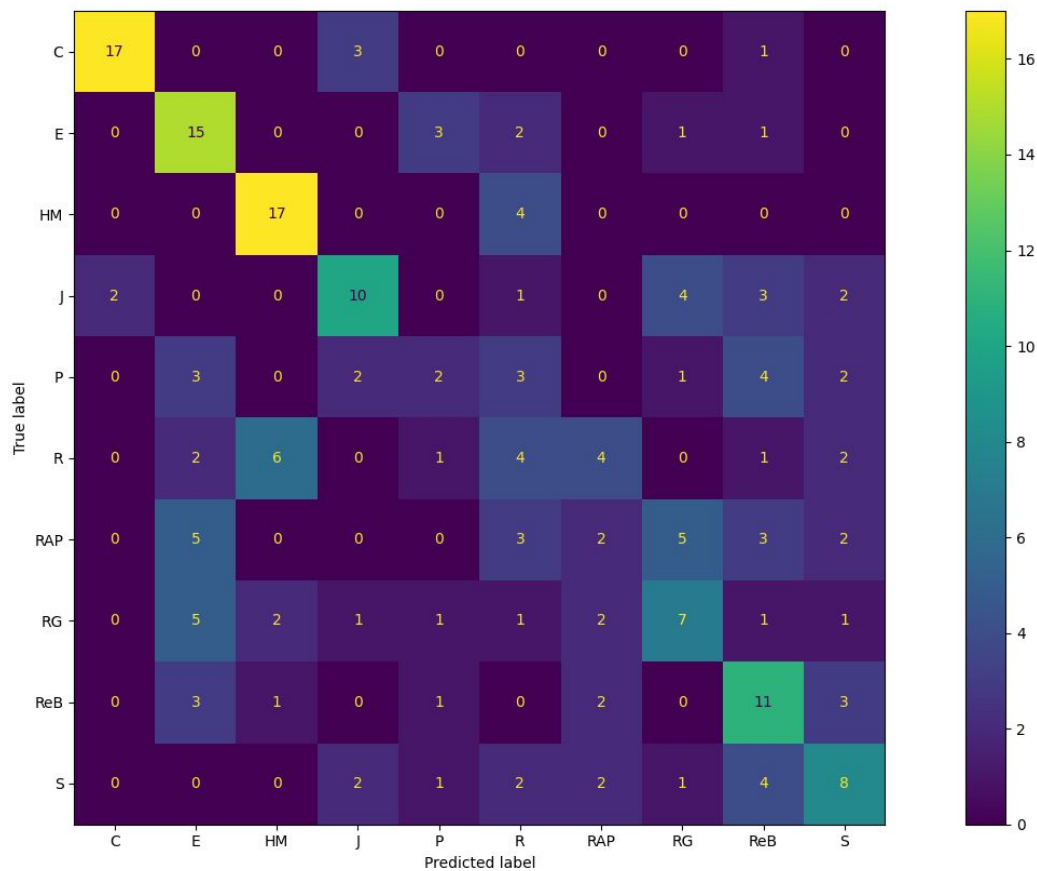
Os resultados foram analisados usando matrizes de confusão.

Resultados



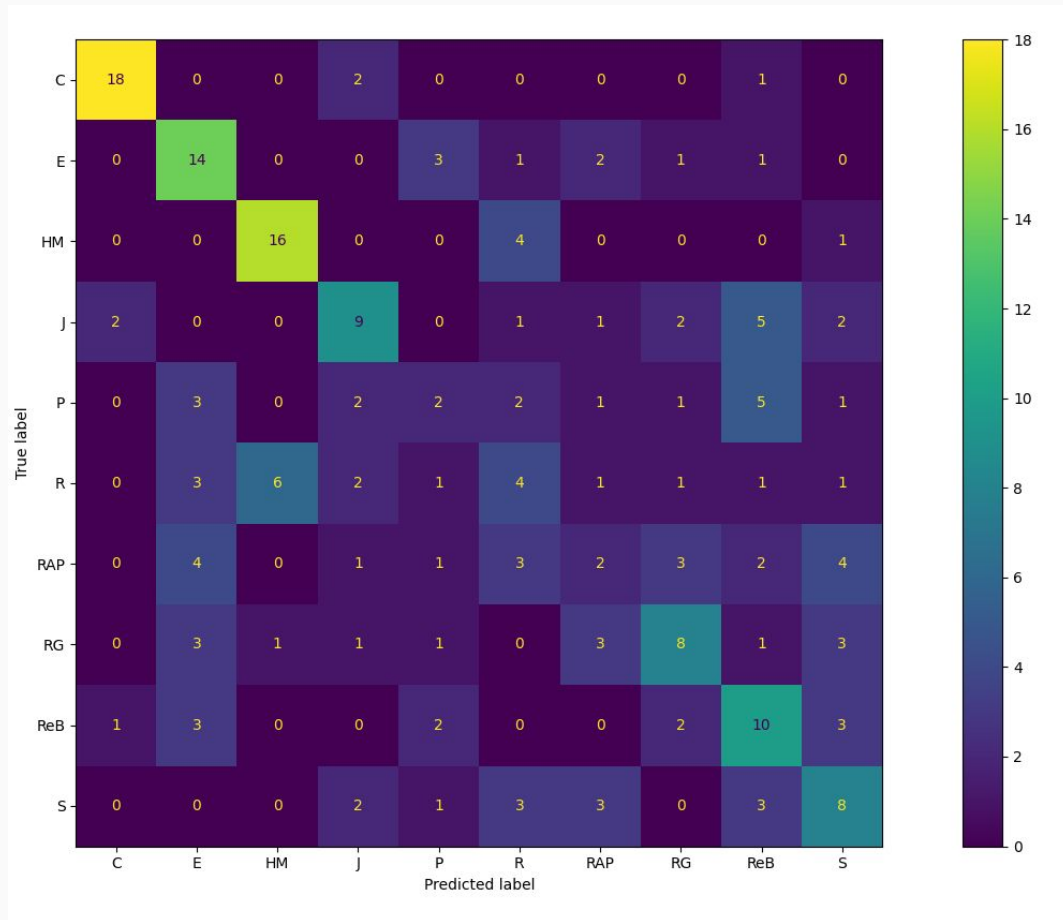
- função de ativação: identidade
- duas camadas ocultas com 200 perceptrons cada
- taxa inicial de aprendizado: 0,01
- algoritmo de solução: adam
- número máximo de iterações: 200
- aleatorização da ordem das entradas
- momentum: 0.9

melhor taxa de acerto: 45,3%



- função de ativação: função hiperbólica tangencial
- uma camada ocultas com 100 perceptrons cada
- taxa inicial de aprendizado: 0,05
- algoritmo de solução: sgd
- número máximo de iterações: 200
- aleatorização da ordem das entradas
- momentum: 0.9

melhor taxa de acerto: 45,3%



- função de ativação: função hiperbólica tangencial
- uma camada oculta com 9 perceptrons
- taxa inicial de aprendizado: 0,1
- algoritmo de solução: descida gradiente estocástica
- número máximo de iterações: 200
- momentum: 0.9

taxa de acerto de 44,3%

Conclusão

Conclusão

1. Uma base de dados maior irá beneficiar em grande escala a predição dos gêneros.
2. Nota-se uma taxa de acerto grande quando usa-se apenas uma camada oculta com 9 perceptrons, o que se aproxima da quantidade de classes sendo analisadas.
3. A rede neural consistentemente agrupou em suas predições Heavy metal e Rock, Clássico e Jazz, Eletrônica e Pop; sugerindo uma aproximação na composição dessas músicas.
4. A rede neural teve consistentemente mais dificuldade, principalmente, com os gêneros Rock, Pop e Rap; sugerindo então que estes são mais ecléticos ou que as músicas selecionadas para representar esse gênero neste banco não são “distintas” o suficiente.
5. Dificuldades gerais com Retrô Gaming, Samba e o R&B

Referências

- https://www.google.com/url?sa=i&url=https%3A%2F%2Fwww.nintendo.pt%2FJogos%2FNES%2FSuper-Mario-Bros-803853.html&psig=AOvVaw1HjsqN9mRgP83wMSZ_klqL&ust=1638471068523000&source=images&cd=vfe&ved=0CAsQjRxqFwoTCNjymsWiw_QCFQAAAAAdAAAAABAD
- https://www.google.com/url?sa=i&url=https%3A%2F%2Fwww.nsctotal.com.br%2Fnoticias%2Fmichael-jackson-esta-vivo-medium-brasileira-afirma-que-sim&psig=AOvVaw2fd53HBsV80VI5BEZBrONQ&ust=1638471257033000&source=images&cd=vfe&ved=0CAsQjRxqFwoTCKiFhp-jw_QCFQAAAAAdAAAAABAD
- https://www.google.com/url?sa=i&url=https%3A%2F%2Fpt.wikipedia.org%2Fwiki%2FClaude_Debussy&psig=AOvVaw0o90q10sSde7qGv95NgrLt&ust=1638471335191000&source=images&cd=vfe&ved=0CAsQjRxqFwoTCNCi8sWjw_QCFQAAAAAdAAAAABAD
- https://en.wikipedia.org/wiki/Spectral_density#Power_spectral_density
- <https://www.kaggle.com/ashishpatel26/feature-extraction-from-audio>
- https://en.wikipedia.org/wiki/Mel-frequency_cepstrum
- <https://scikit-learn.org/>

Obrigado por
assistir!

:D