

Aprendizado de Máquina e Classificação de Músicas

Luciano Pinheiro Batista

{lucianopinheiro15@gmail.com}

Faculdade de Engenharia Elétrica e de Computação (FEEC)

Universidade Estadual de Campinas (Unicamp)

Campinas, SP, Brasil

Resumo – Este artigo tem como objetivo fornecer métodos de classificação de músicas utilizando técnicas de Aprendizado de Máquina Supervisionado e Não Supervisionado. Em ambos os casos um conjunto de características musicais foi recolhido com o auxílio da *Web API* do Spotify [5]. No caso supervisionado, os dados foram utilizados para o treinamento de uma rede *Multilayer Perceptron*. A rede foi capaz de reconhecer padrões de teste de 6 gêneros diferentes com uma acurácia de 87.8%. No caso não supervisionado, o algoritmo *k-means* foi aplicado aos dados e conjuntos de músicas com características semelhantes foram agrupados em clusters.

Palavras-chave – Aprendizado Supervisionado, Aprendizado Não Supervisionado, Multilayer Perceptron, Clusterização, Classificação de Músicas.

1. Introdução

Os novos meios de Distribuição Eletrônica de Música e a exploração sistemática de grandes bases de dados musicais criam uma necessidade de produzir descrições simbólicas para músicas [7]. Neste cenário, o gênero musical é provavelmente a forma mais amplamente utilizada de descrição musical. Apesar de ser um conceito de difícil definição, o gênero está intrinsecamente relacionado a tarefa de classificação musical. Designar um gênero a uma música é uma maneira útil de descrever o que essa música tem em comum com outras do mesmo gênero e o que a torna diferente de músicas de outros gêneros [7].

No contexto de Aprendizado de Máquina Supervisionado a designação de gêneros musicais pode ser estruturada em termos de um problema de classificação. Este problema consiste em atribuir a cada padrão de entrada o rótulo correspondente a uma das Q classes existentes, C_k , $k = 1, \dots, Q$, à qual o dado pertence [8]. No presente trabalho, os padrões de entrada serão características de músicas recolhidas a partir da ferramenta *Get Audio Analysis* da *web API* do Spotify [5]. Estas características fornecem, portanto, uma representação numérica de cada música. As classes associadas a cada padrão, por sua vez, serão um dos seguintes 6 gêneros musicais: *Techno*, *Reggae*, *Samba*, *Punk*, *Metal* e *Rap*. A partir destes dados uma rede neural do tipo *Multilayer Perceptron* (MLP), foi treinada com o objetivo de classificar corretamente uma música em um determinado gênero.

Uma vez que nem sempre é fácil rotular uma música com um gênero em específico [7], a abordagem não supervisionada pode ser útil para realizar o agrupamento de músicas com características semelhantes. Neste trabalho isto será realizado por meio de uma clusterização.

A clusterização tem como objetivo identificar padrões semelhantes e atribuí-los a grupos de padrões semelhantes [9]. Os padrões foram recolhidos assim como no caso supervisionado e separados em clusters por meio do algoritmo *k-means*. Os clusters obtidos podem ser utilizados, por exemplo, para identificar qual a melhor maneira de separar um conjunto de músicas para a criação de um certo número de playlists com músicas semelhantes.

2. Metodologia

2.1. Aquisição dos Dados

Para que os métodos de aprendizado de máquina possam ser aplicados à tarefa de classificação musical é necessário que as músicas sejam representadas por meio de números. Para isso, dados numéricos de cada música foram recolhidos por meio da *Web API* do Spotify [5]. Utilizando a biblioteca *Spotipy* [6] os dados foram acessados de maneira simples usando Python. Com o auxílio da operação *Get-Audio-Analysis* [5] os seguintes atributos foram extraídos de cada música:

- Duração
- Volume
- Andamento
- Tempo
- Tom
- Média da Duração de Tatus
- Média da Duração de Segmentos
- Média da Duração de Compassos
- Média da Duração de Seções
- Média da Duração de Batidas
- Média do Timbre dos Segmentos
- Média da Altura dos Segmentos

Desse modo cada música passou a ser representada por um vetor pertencente ao \mathbb{R}^{12} . Um detalhamento dos conceitos envolvidos nos itens acima foge do escopo deste artigo, mas pode ser encontrado em referências de teoria musical como [11].

2.2. Aprendizado Supervisionado

A tabela 1 mostra os gêneros e os rótulos numéricos associados a cada um deles. Uma busca no Spotify foi realizada para cada um dos gêneros musicais e as duas primeiras playlists encontradas para cada gênero foram selecionadas. A partir das músicas contidas nas playlists selecionadas uma base de dados de 817 músicas foi construída com o cuidado de evitar que houvesse grande desbalanceamento entre as classes. Os atributos de cada música foram extraídos conforme a seção 2.1. e cada uma recebeu um rótulo numérico de acordo com o gênero da playlist a qual foi retirada.

Gênero	Rótulo	Nº de Padrões
Techno	0	160
Reggae	1	100
Samba	2	92
Punk	3	100
Metal	4	150
Clássica	5	215

Tabela 1. Rótulos atribuídos a cada gênero e Quantidade de Padrões por Gênero.

As características de cada padrão foram normalizadas utilizando o máximo valor encontrado para cada característica, a fim de evitar que certos atributos se tornassem dominantes em relação aos demais durante o processo de treinamento. Após isso, os dados foram separados em um conjunto de treinamento - 2/3 do total - e outro de validação - 1/3 do total. Uma rede MLP criada com o auxílio da biblioteca scikit-learn [10] foi, então, alimentada com o conjunto de treinamento. O número de camadas, o número de neurônios por camada, a função de ativação, a taxa de aprendizagem e o método de otimização foram escolhidos por meio de validação cruzada com base na acurácia obtida junto aos dados de validação.

Uma vez validada, a rede foi posta à prova para classificar músicas de álbuns considerados clássicos de cada gênero segundo publicações especializadas como [3] e [4].

2.3. Aprendizado Não Supervisionado

Para a abordagem de aprendizado não supervisionado foi realizada uma busca no Spotify com as palavras *Happy* e *Sad*. As duas primeiras playlists encontradas para cada

busca foram selecionadas e um banco de dados de 389 músicas foi construído com os atributos extraídos de cada música. Novamente utilizando a biblioteca scikit-learn, o algoritmo *k-means* foi aplicado aos dados com diferentes valores do parâmetro *k*, que nada mais é que o número de clusters escolhido. Após isso, o *Silhouette Score* de cada separação foi calculado com o objetivo de identificar o melhor número de clusters que o conjunto de músicas pode ser agrupado.

Em contraste com a seção 2.2. os rótulos das músicas não foram utilizados no processo de separação, uma vez que no cenário não supervisionado os dados de treinamento não são rotulados, em vez disso o sistema tenta aprender sem um professor [9].

3. Resultados e Análise

3.1. Aprendizado Supervisionado

Não foi notada uma grande diferença de resultados entre as funções de ativação e o método de otimização utilizados junto aos dados de validação. Contudo a combinação Gradiente Descendente Estocástico e Tangente Hiperbólica possibilitaram ao modelo uma melhor capacidade de generalização ao se aplicar os álbuns de teste da tabela 3.

A figura 1 mostra a evolução das acurácias obtidas com essa configuração junto aos dados de validação ao utilizar duas camadas com o mesmo número de neurônios e uma taxa de aprendizado 0.01. Percebe-se que a partir de 12 neurônios por camada a acurácia permanece em torno de 85%.

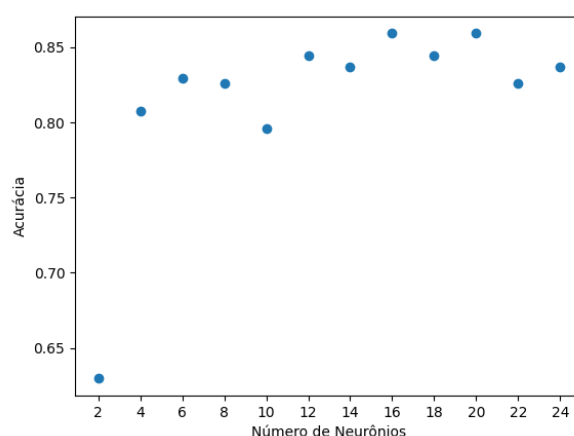


Figura 1. Acurácia vs. Nº de Neurônios

A tabela 2 mostra a matriz de confusão obtida ao alimentar com os dados de validação uma rede com duas camadas de 20 neurônios. Com esta configuração foi possível obter uma acurácia de 87.8%. Contudo, é importante ressaltar que a rede se mostrou sensível a inicialização dos pesos sinápticos e houve, em função disso, uma flutuação da acurácia obtida para uma mesma configuração de rede treinada.

É interessante notar também que os gêneros Punk e Metal, por serem ambos sub-gêneros do Rock, e consequentemente apresentarem um maior número de características em comum, confundiram levemente a rede. Apesar disso a taxa de verdadeiro positivo permaneceu acima dos 85% em todas as classes, exceto a 2 - Samba. Isto pode ter ocorrido porque esta classe teve menos padrões recolhidos em relação às outras, prejudicando o seu reconhecimento pela rede.

	0	1	2	3	4	5
Techno	49	4	3	0	1	1
Reggae	0	23	4	0	0	0
Samba	0	8	20	1	0	1
Punk	0	0	0	29	3	0
Metal	0	0	0	6	43	0
Clássica	0	0	1	0	0	73

Tabela 2. Matriz de Confusão

A taxa de acertos da rede quando alimentada com músicas de álbuns considerados clássicos de cada gênero pode ser observada na tabela 3. O único trabalho com uma taxa de acertos abaixo de 80% foi o álbum Selected Ambient Works. Isto se deve provavelmente ao caráter mais experimental do autor deste projeto [1], o que levou a rede a não classificá-lo tão corretamente quanto os outros.

Álbum	Rótulo	% de Acertos
Selected Ambient Works	0	70%
Exodus	1	83%
Cartola (1976)	2	85.7%
Never Mind The Bollocks	3	81%
Master of Puppets	4	100%
Nocturnes Vol. 1	5	100%

Tabela 3. Taxa de Acertos em Álbuns Clássicos

3.2. Aprendizado Não Supervisionado

A figura 2 mostra a evolução do *Silhouette Score* em relação ao aumento do número de clusters gerados pelo algoritmo *k-means*. A melhor separação, obtida quando o *Silhouette Score* é o mais alto, ocorre com $K = 2$, ou

seja, quando o conjunto de músicas é separado em dois grupos.

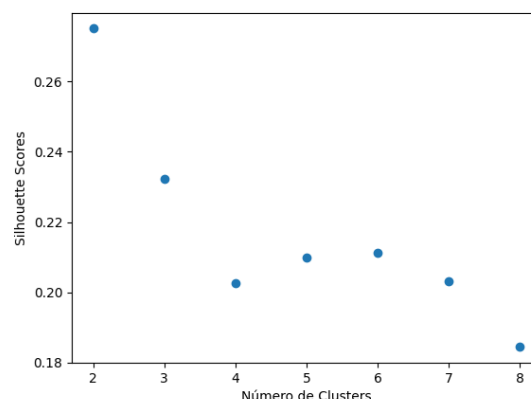


Figura 2. Silhouette Scores vs. Nº de Clusters

Este comportamento pode ter sido influenciado pelo método como as músicas em questão foram obtidas na seção 2.3.. De fato, a tabela 4 mostra que houve uma correlação entre os clusters e a porcentagem de músicas deste cluster provenientes de cada uma das buscas realizadas.

Cluster	% busca "Happy"	% busca "Sad"
0	79%	21%
1	19%	81%

Tabela 4. Clusters e Procedência da Música

De certa forma, isso é um indicativo de que o cluster foi capaz de separar músicas com características semelhantes, ainda que não tenha sido atribuído um rótulo às músicas no processo de separação. Em um cenário de seleção aleatória de músicas, este método pode ser eficaz para escolher a melhor maneira de agrupar músicas com características semelhantes em diferentes playlists.

4. Conclusão e Perspectivas Futuras

Neste trabalho foi apresentado como o aprendizado de máquina pode ser aplicado a tarefa de classificação musical. Duas abordagens diferentes foram propostas, uma baseada em aprendizado supervisionado e a outra em aprendizado não supervisionado. A primeira se mostrou interessante em classificar músicas em diferentes gêneros, enquanto a segunda foi eficiente em agrupar músicas com características em comum. O código utilizado para a obtenção dos resultados pode ser encontrado em [2].

Para a primeira abordagem um desafio seria aumentar o número de gêneros, isto é, de classes possíveis e verificar se a rede MLP mantém uma boa acurácia. A

medida em que mais dados seriam necessários, maiores seriam as chances de recolher padrões que não se enquadram perfeitamente em um determinado gênero, ou de haver um gênero que incorpora elementos de todos os outros gêneros, dificultando a sua classificação.

Uma técnica que pode ser explorada conjuntamente com o que foi apresentado neste artigo é o processamento de linguagem natural. Ele pode ser utilizado de maneira conjunta para incorporar ao modelo características das letras das músicas. Isto pode eventualmente ajudar na tarefa de distinguir diferentes gêneros, ou agrupar músicas com base em atributos semelhantes.

Referências

- [1] Aphex Twin Wikipedia Page. https://en.wikipedia.org/wiki/Aphex_Twin. Acessado: 28-11-2022.
- [2] Aprendizado de Máquina e Classificação Musical. https://github.com/lucianopinheirob/Classificacao_Musical. Acessado: 30-11-2022.
- [3] Pitchfork Website. <https://pitchfork.com/>. Acessado: 28-11-2022.
- [4] Rolling Stone Website. <https://rollingstone.uol.com.br/>. Acessado: 28-11-2022.
- [5] Spotify Web API. <https://developer.spotify.com/documentation/web-api/>. Acessado: 28-11-2022.
- [6] Spotipy, (2022), A light weight Python library for the Spotify Web API. <https://github.com/plamere/spotipy>. Acessado: 28-11-2022.
- [7] Jean-Julien Aucouturier and Francois Pachet. Representing musical genre: A state of the art. *Journal of new music research*, 32(1):83–93, 2003.
- [8] Richard O. Duda, Peter E. Hart, and David G. Stork. *Pattern Classification*. Wiley, New York, 2 edition, 2001.
- [9] Aurelien Geron. *Hands-on machine learning with Scikit-Learn and TensorFlow : concepts, tools, and techniques to build intelligent systems*. O'Reilly Media, Sebastopol, CA, 2017.
- [10] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, and E. Duchesnay. Scikit-learn: Machine learning in Python. *Journal of Machine Learning Research*, 12:2825–2830, 2011.
- [11] Don Michael Randel. *The Harvard dictionary of music*. Harvard University Press, 2003.