Aplicação da rede SOM para identificação de gênero musical

Flávio B. S. Mota[1], **Marcos G. Quiles**[1]

[1]Instituto Nacional de Pesquisas Espaciais (INPE)
Avenida dos Astronautas – 1758 – Jardim da Granja – São José dos Campos
SP – 12227-010 – Brasil

flavio.belizario.mota@gmail.com, marcos.quiles@inpe.br

Abstract. *Several applications employ the SOM neural network method as a way of organizing data. Music data sets can be organized considering the genre of each song. There are several features that can be considered to define the genre of a song. This work analyzes a set of attributes extracted from songs, using the SOM network as a technique to group and identify a musical genre. Rock and reggae genres were specifically analyzed and the results indicate that the set of attributes from the songs' audio produced good clusters, although not sufficiently adequate for the task.*

Resumo. *Diversas aplicações empregam o método de rede neural não supervisionada SOM como forma de organizar dados. Conjuntos de dados musicais podem ser organizados considerando o estilo de cada canção. Existem vários atributos que podem ser considerados para definir o gênero de uma música. Esse trabalho analisa um conjunto de atributos extraído de músicas, empregando a rede SOM como técnica para agrupar e identificar um gênero musical. Foram analisados especificamente os gêneros rock e reggae e os resultados apontam que o conjunto de atributos provenientes dos áudios das canções produziram bons agrupamentos, ainda que não suficientemente adequados para a tarefa.*

1. Introdução

Redes SOM (*Self-Organizing Maps*), ou Mapas Auto-Organizáveis, são tipos de redes neurais fortemente associadas à percepção humana, por conta de sua fundamentação neurofisiológica [Kohonen 1989]. Seu uso pode ser percebido em aplicações voltadas para a exploração de bases de dados [Honkela et al. 1996], diagnósticos financeiros [Serrano-Cinca 1996], controle de processos químicos [Ultsch 1993], garimpagem de dados [Ong and Abidi 1999], entre outras. Muitas dessas aplicações visam organizar os dados por sua similaridade [Braga et al. 2007].

Músicas geralmente são organizadas em função de diferentes gêneros como rock, pop, country, blues etc. Esses gêneros diferem não apenas no som, no humor e na história, mas também no público-alvo [Ahmad et al. 2014]. Essa organização dos gêneros musicais é considerada um tanto quanto subjetiva, pois existem músicas que podem ser entendidas como pertencentes a vários gêneros, ou algumas que não se enquadram em nenhum dos gêneros já identificados [Tao Li and Tzanetakis 2003]. No entanto, a organização de forma automática das músicas por gêneros pode ser uma aplicação de muita utilidade considerando estações de rádio, gerenciamento de arquivos audiovisuais da internet, entretenimento e outros.

Embora seja difícil definir com exatidão quais são os atributos específicos que caracterizam um gênero musical, diversos trabalhos se propõem a analisar e extrair atributos das músicas. O objetivo desse trabalho é apresentar a análise do emprego da rede SOM na tarefa de agrupamento (organização) de músicas de dois gêneros: rock e reggae, considerando os atributos disponibilizados na base de dados fruto do estudo de [Misael et al. 2020].

2. Revisão da Literatura

Desenvolvidas em 1982 por Teuvo Kohonen [Kohonen 1982], as redes SOM (do inglês *Self-Organizing Maps*, Mapas Auto-Organizáveis) são fruto do estudo de estruturas neurofisiológicas presentes do córtex cerebral. De forma geral, o cérebro possui áreas responsáveis por funções específicas como a fala, visão, controle motor. Dentro dessas áreas, os neurônios possuem uma ordenação espacial, sendo que neurônios topologicamente próximos respondem por estímulos e padrões semelhantes. Sendo assim, o funcionamento da rede consiste em receber um padrão de entrada **p** e encontrar um neurônio que seja mais parecido com esse padrão. Durante o treinamento, a semelhança entre o neurônio escolhido é aumentada, assim como a de seus vizinhos [Kohonen 1997].

A rede SOM é um tipo de rede neural de camada única na qual os neurônios estão distribuídos em uma grade de n dimensões. Para a etapa de treinamento da rede, é empregado o algoritmo de aprendizado competitivo, no qual os neurônios competem entre si para se tornarem ativos, sendo que aquele que se aproxima mais do padrão de entrada apresentado vence [Richardson et al. 2003]. Para decidir qual dos neurônios é o vencedor, é empregada uma medida de distância entre o padrão de entrada apresentado (x) e os pesos dos neurônios (w). Geralmente a medida empregada é a distância euclidiana [Miljkovic 2017], dada pela Equação 1.

$$d_j = \sum_{i=1}^{n} (x_i - w_{ij})^2 \tag{1}$$

Depois da escolha do neurônio vencedor, acontece a atualização dos pesos desse neurônio e dos neurônios vizinhos, limitada por um raio de vizinhança, de forma a garantir a ordenação topológica da rede [Braga et al. 2007]. A Equação 2 apresenta a atualização de pesos do neurônio vencedor e de sua vizinhança:

$$w_{ji}(t+1) = \begin{cases} w_{ji}(t) + \eta(t)(x_i(t) - w_{ji}(t)), & se \;\; j \in \Lambda(t) \\ w_{ji}(t), & caso \;\; contrário \end{cases} \tag{2}$$

onde $w_{ji}(t)$ é o peso da conexão entre o neurônio j e o vetor de entrada $x_i(t)$, $\eta(t)$ é a taxa de aprendizado e $\Lambda(t)$ é a vizinhança, todos no tempo t. Durante o treinamento, tanto a taxa de aprendizado quanto o raio da vizinhança são continuamente decrementados, visando garantir convergência e estabilidade.

O modelo SOM está presente em diferentes aplicações com dados sobre músicas. No trabalho de [Rauber et al. 2002] a rede foi utilizada como estratégia para organizar uma biblioteca musical de acordo com os gêneros musicais. Também no intuito de organizar um banco de dados de músicas, [Frühwirth and Rauber 2002] utilizaram características baseadas em conteúdo e gênero das músicas como entrada da rede SOM. Em

[Palomäki et al. 1999] os autores aplicaram a rede SOM para a avaliação da discriminação espacial de fontes sonoras reais e virtuais, simulando a percepção humana do som espacial. O trabalho de [Tuzman 2001] empregou a técnica para restauração e redução de ruídos em músicas. Segundo o trabalho de [Plewa and Kostek 2015], a aplicação mais comum das redes SOM no contexto de análise de músicas é para a geração de representações bidimensionais de conjuntos, bases de dados ou amostras específicas de canções.

Nos trabalhos citados anteriormente, existem diversos atributos que são utilizados para descrever as músicas, como atributos extraídos diretamente da análise dos sinais de áudio ou provenientes da análise do conteúdo lírico. O trabalho de [Misael et al. 2020] propõe um método para extração de atributos de um conjunto de músicas empregando *Latent Dirichlet Allocation* para analisar o conteúdo das letras. Através dos resultados obtidos, foram modelados tópicos que passaram a compor o conjunto de atributos que descrevem as músicas analisadas. O propósito dessa abordagem foi de investigar e mostrar a evolução musical dos gêneros musicais mais comuns considerando os tópicos, mas além disso resultou na elaboração de uma nova base de dados, com novos atributos. O conjunto de dados será abordado com mais detalhes na próxima Seção.

3. Materiais e Métodos

Os dados empregados nesse trabalho são provenientes da base de dados resultado do estudo *Temporal Analysis and Visualisation of Music* [Misael et al. 2020]. Para construção da base, os autores coletaram músicas de 7 gêneros (rock, reggae, jazz, blues, hip hop, country, pop) e para cada canção coletada, foram extraídos atributos fornecidos pelo serviço de *streaming* Spotify, sendo eles: *Acousticness*, *Danceability*, *Loudness*, *Instrumentalness*, *Valence*, *Energy*. Esses atributos foram considerados nesse estudo como atributos musicais, por serem resultado da análise do aúdio da canção.

Além desses atributos, através da aplicação da técnica de *Latent Dirichlet Allocation* (LDA) nas letras das canções, novos atributos baseados nos tópicos encontrados foram adicionados ao conjunto de dados. Os tópicos gerados foram: *dating*, *violence*, *world/life*, *night/time*, *shake the audience*, *family/gospel*, *romantic*, *communication/thinking*, *obscene*, *sound*, *movement/places*, *light/visual perceptions*, *family/spiritual*, *like/girls*, *sadness*, *feelings*. Ao final, os autores produziram um conjunto contendo 28372 amostras e 30 atributos. A distribuição das amostras por gênero é apresentada na Tabela 1.

Tabela 1. Quantidade de amostras por gênero musical.

Gênero	Amostras
pop	7042
country	5445
blues	4604
rock	4034
jazz	3845
reggae	2498
hip hop	904

Nesse trabalho, foram elaboradas 3 análises considerando o conjunto de dados

descrito, no intuito de agrupar as músicas por sua similaridade. A primeira análise considera o agrupamento empregando os atributos provenientes do serviço de *streaming* Spotify, a segunda análise leva em consideração os atributos gerados pelo LDA e por fim uma análise com todos os atributos disponibilizados. Além disso, foram utilizadas apenas as amostras dos gêneros rock e reggae. Isso se deve ao fato de que, na análise de correlação entre gêneros conduzida no trabalho de [Misael et al. 2020], rock e reggae são os que apresentam menor similaridade.

Para os experimentos, foi utilizada a implementação da rede SOM disponibilizada na biblioteca MiniSom da linguagem Python [Vettigli 2018]. Durante os experimentos, foram avaliadas diferentes arquiteturas de rede, variando o tamanho da grade e mantendo os valores padrão de vizinhança (*sigma*) e taxa de aprendizado da biblioteca, respectivamente 1.0 e 0.5. Foram observados os valores dos erros de quantização e topologia da rede. Após encontrar o tamanho de grade que minimizou ambos os erros, foram testados diferentes valores de vizinhança e taxa de aprendizagem, buscando minimizar ainda mais os valores de erro. Os experimentos foram executados em uma máquina com processador *Intel®Core™i5-10500H CPU @ 2.50GHz* e 8*GB* de *RAM*. Os resultados obtidos são descritos e discutidos na Seção a seguir.

4. Resultados

Para avaliar os resultados obtidos, foram considerados os valores dos erros de quantização e topologia da rede. Variando o tamanho da rede, verificou-se que a grade de 25x25 obteve os menores valores de erro de quantização e topografia. Valores de grade acima de 25x25 não apresentaram diminuição significativa dos valores de erro. Sendo assim, os parâmetros de vizinhança (*sigma*) e taxa de aprendizado dessa configuração de grade foram variados até atingir-se os menores erros. Todos as configurações foram executadas por 10000 épocas. As Tabelas 2, 3 e 4 apresentam os resultados obtidos com as variações de tamanho de grade e os valores de parâmetros da rede de grade 25x25 que apresentaram melhor resultado.

Tabela 2. Resultados da rede SOM com atributos musicais.

Atributos musicais				
Grade	Sigma	Taxa de Aprendizado	Quantização	Topologia
5x5	1.0	0.5	0.22	0.21
10x10	1.0	0.5	0.15	0.35
25x25	1.0	0.5	0.12	0.32
25x25	1.8	0.5	0.11	0.12

Tabela 3. Resultados da rede SOM com atributos baseados em tópicos.

Atributos baseados em tópicos				
Grade	Sigma	Taxa de Aprendizado	Quantização	Topologia
5x5	1.0	0.5	0.23	0.30
10x10	1.0	0.5	0.18	0.36
25x25	1.0	0.5	0.16	0.37
25x25	2.1	0.5	0.15	0.13

Tabela 4. Resultados da rede SOM com combinação de atributos.

Combinação de atributos				
Grade	Sigma	Taxa de Aprendizado	Quantização	Topologia
5x5	1.0	0.5	0.44	0.50
10x10	1.0	0.5	0.37	0.42
25x25	1.0	0.5	0.32	0.41
25x25	2.0	0.3	0.32	0.12

É possível notar que os menores valores de erro foram obtidos quando empregados apenas os atributos provenientes das características musicais. Ainda em relação aos atributos musicais, se comparados aos resultados obtidos utilizando a combinação de atributos, o erro de topologia da rede é o mesmo, porém o erro de quantização aumenta. Isso pode indicar que os atributos musicais são mais adequados para realizar o agrupamento das músicas.

Como o conjunto de dados possui os gêneros de cada amostra, é possível visualizar a matriz de distâncias dos neurônios da rede com os rótulos de cada gênero associado ao neurônio correspondente, como apresenta a Figura 1. Na figura, quadrados azuis representam o gênero reggae e os 'x' vermelhos representam o gênero rock. A tonalidade de fundo representa a distância de um neurônio em relação aos seus vizinhos. A partir da análise dessa imagem é possível notar que, ainda que existam neurônios que se especializaram em determinados grupos, muitos neurônios agrupam amostras de ambos os gêneros. Isso indica que, mesmo com valores baixos de erro de quantização e topologia, a rede tem dificuldade em separar completamente os gêneros, ainda que existam regiões no mapa bem definidas.

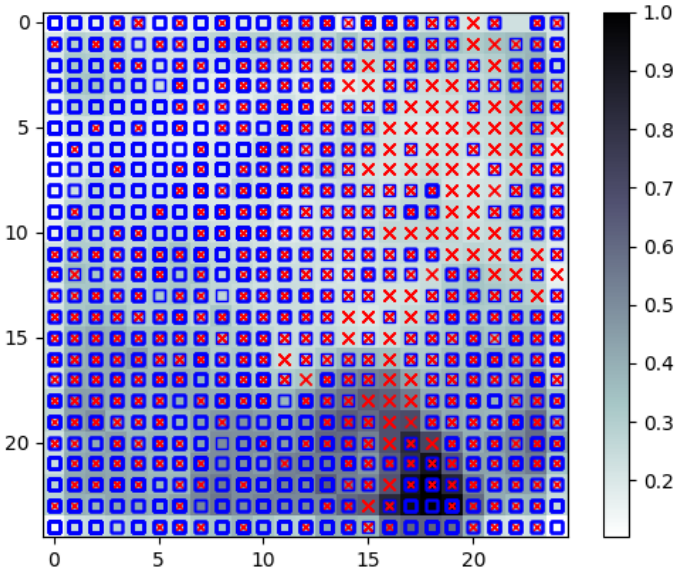

Figura 1. Matriz de distâncias da rede SOM com os rótulos (gêneros). Quadrados azuis representam o gênero reggae e 'x' vermelhos representam o gênero rock.

5. Conclusões

Nesse trabalho foi analisada a aplicação de uma rede SOM na tarefa de agrupamento de músicas considerando a diferença entre dois gêneros musicais. Foram comparados diferentes conjuntos de atributos, afim de avaliar quais são mais adequados para a organização das amostras. Os resultados apontaram que empregar o conjunto de atributos provenientes da análise musical produziu menores erros de quantização e topologia, empregando uma arquitetura de rede com uma grade de 25x25, *sigma* de 1.8 e taxa de aprendizagem de 0.5. Analisando os agrupamentos gerados, foi possível observar que, ainda que muitos neurônios consigam se especializar em um grupo específico, muitos agrupam amostras de ambos os gêneros. Isso aponta que, mesmo que se baseando-se nos valores dos atributos musicais, é difícil separar completamente as amostras dos gêneros rock e reggae, indicando que esses atributos não são suficientemente adequados. Cabe teorizar que tais estilos compartilham muitas características em relação à musicalidade e, por isso, é mais complicado diferenciar os gêneros.

Como trabalhos futuros, sugere-se o estudo de arquiteturas de redes com mais de duas dimensões no intuito de obter uma melhor separação do conjunto. Além disso, é possível também o estudo de uma rede SOM *Fuzzy*, como proposto em [Vuorimaa 1994], afim de analisar o grau de pertencimento das amostras associadas à cada grupo, possibilitando assim que neurônios que agrupam amostras de mais de um gênero possam ser analisados considerando músicas que compartilham atributos musicais.

Referências

Ahmad, A. N., Sekhar, C., and Yashkar, A. (2014). Music Genre Classification Using Music Information Retrieval and Self Organizing Maps. In Pant, M., Deep, K., Nagar, A., and Bansal, J. C., editors, *Proceedings of the Third International Conference on Soft Computing for Problem Solving*, volume 259, pages 625–634. Springer India, New Delhi. Series Title: Advances in Intelligent Systems and Computing.

Braga, A. d. P., Ludermir, T. B., and Carvalho, A. C. P. d. L. F. (2007). *Redes neurais artificiais: teoria e aplicações*. LTC, Rio de Janeiro, 1 edition.

Frühwirth, M. and Rauber, A. (2002). Self-Organizing Maps for Content-Based Music Clustering. In Taylor, J. G., Tagliaferri, R., and Marinaro, M., editors, *Neural Nets WIRN Vietri-01*, pages 228–233. Springer London, London. Series Title: Perspectives in Neural Computing.

Honkela, T., Kaski, S., Lagus, K., and Kohonen, T. (1996). Exploration of full-text databases with self-organizing maps. In *Proceedings of International Conference on Neural Networks (ICNN'96)*, volume 1, pages 56–61, Washington, DC, USA. IEEE.

Kohonen, T. (1982). Self-organized formation of topologically correct feature maps. *Biological Cybernetics*, 43(1):59–69.

Kohonen, T. (1989). *Self-Organization and Associative Memory*, volume 8 of *Springer Series in Information Sciences*. Springer Berlin Heidelberg, Berlin, Heidelberg.

Kohonen, T. (1997). *Self-Organizing Maps*, volume 30 of *Springer Series in Information Sciences*. Springer Berlin Heidelberg, Berlin, Heidelberg.

Miljkovic, D. (2017). Brief review of self-organizing maps. In *2017 40th International Convention on Information and Communication Technology, Electronics and Microelectronics (MIPRO)*, pages 1061–1066, Opatija, Croatia. IEEE.

Misael, L., Forster, C., Fontelles, E., Sampaio, V., and França, M. (2020). Temporal Analysis and Visualisation of Music. In *Anais do Encontro Nacional de Inteligência Artificial e Computacional (ENIAC 2020)*, pages 507–518, Brasil. Sociedade Brasileira de Computação - SBC.

Ong, J. and Abidi, S. S. R. (1999). Data mining using self-organizing kohonen maps: A technique for effective data clustering & visualization. In Arabnia, H. R., editor, *Proceedings of the International Conference on Artificial Intelligence, IC-AI '99, June 28 - July 1, 1999, Las Vegas, Nevada, USA, Volume 1*, pages 261–264. CSREA Press.

Palomäki, K., Pulkki, V., and Karjalainen, M. (1999). Neural network approach to analyze spatial sound. *Journal of the Audio Engineering Society*.

Plewa, M. and Kostek, B. (2015). Music Mood Visualization Using Self-Organizing Maps. *Archives of Acoustics*, 40(4):513–525.

Rauber, A., Pampalk, E., and Merkl, D. (2002). Content-based music indexing and organization. In *Proceedings of the 25th annual international ACM SIGIR conference on Research and development in information retrieval - SIGIR '02*, page 409, Tampere, Finland. ACM Press.

Richardson, A., Risien, C., and Shillington, F. (2003). Using self-organizing maps to identify patterns in satellite imagery. *Progress in Oceanography*, 59(2-3):223–239.

Serrano-Cinca, C. (1996). Self organizing neural networks for financial diagnosis. *Decision Support Systems*, 17(3):227–238.

Tao Li and Tzanetakis, G. (2003). Factors in automatic musical genre classification of audio signals. In *2003 IEEE Workshop on Applications of Signal Processing to Audio and Acoustics (IEEE Cat. No.03TH8684)*, pages 143–146, New Paltz, NY, USA. IEEE.

Tuzman, A. (2001). Wavelet and self-organizing map based declacker. *Journal of the Audio Engineering Society*.

Ultsch, A. (1993). Self Organized Feature Maps for Monitoring and Knowledge Aquisition of a Chemical Process. In Gielen, S. and Kappen, B., editors, *ICANN '93*, pages 864–867. Springer London, London.

Vettigli, G. (2018). Minisom: minimalistic and numpy-based implementation of the self organizing map.

Vuorimaa, P. (1994). Fuzzy self-organizing map. *Fuzzy Sets and Systems*, 66(2):223–231.