

Musical Genre Clustering With Self Organizing Maps

Giancarlo Klemm

Resumo—Lorem ipsum dolor sit amet, consectetur adipiscing elit. Etiam lobortis facilisis sem. Nullam nec mi et neque pharetra sollicitudin. Praesent imperdiet mi nec ante. Donec ullamcorper, felis non sodales commodo, lectus velit ultrices augue, a dignissim nibh lectus placerat pede. Vivamus nunc nunc, molestie ut, ultricies vel, semper in, velit. Ut porttitor. Praesent in sapien. Lorem ipsum dolor sit amet, consectetur adipiscing elit. Duis fringilla tristique neque. Sed interdum libero ut metus. Pellentesque placerat. Nam rutrum augue a leo. Morbi sed elit sit amet ante lobortis sollicitudin. Praesent blandit blandit mauris. Praesent lectus tellus, aliquet aliquam, luctus a, egestas a, turpis. Mauris lacinia lorem sit amet ipsum. Nunc quis urna dictum turpis accumsan semper.

Keywords—*IEEEtran, journal, L^AT_EX, paper, template.*

I. INTRODUCTION

O clustering de músicas por gênero é um problema complexo. A extração de informações de uma música digitalizada já é extremamente complexa, mas a utilização desses dados para classificar, comparar e fazer processamentos em geral também não é uma tarefa simples. Este paper se dedica a verificar se métodos bio-inspirados geram bons resultados para calcular a similaridade entre músicas, verificando se músicas do mesmo gênero são sempre similares se comparadas com outros gêneros musicais.

O computação bio-inspirada é o uso de computadores para modelar fenômenos biológicos. Pesquisas recentes mostram que modelos bio-inspirados podem ajudar a resolver vários problemas computacionais. Um dos mais relevantes são as redes neurais, modelos que são usados para aproximar funções que dependem de um número grande de entradas e que são representados como um conjunto de neurônios interligados.

Um tipo mais específico de rede neural é o mapa de Kahonen, que tem como saída uma representação de baixa dimensão das entradas. Com um mapa de Kahonen é possível gerar um mapa de similaridade dos dados de entrada. Por isto, este método foi aplicado em um conjunto de músicas para verificar como elas são representadas e qual a relação com o seu gênero.

II. TRABALHO ANTERIOR

Os dados utilizados nestes experimentos são processados utilizando parte do programa descrito em [3]. Este programa utiliza a biblioteca Marsyas¹ para fazer a extração de características de músicas digitalizadas para poder fazer comparações

entre elas. Nesta implementação algumas características já foram testadas para verificar quais são melhores para comparações, logo estas são usadas para fazer o cálculo de similaridade no mapa de Kahonen.

III. MAPA DE KAHONEN

De acordo com [1], o mapa de Kahonen é um tipo de rede neural treinada com métodos não-supervisionados. Ela é capaz de gerar uma representação discreta de baixa dimensão para um conjunto de dados com alta dimensão. Nesta publicação os dados com alta dimensão são as características extraídas de músicas digitalizadas, enquanto que a saída é um gráfico 2D representando a semelhança entre cada elemento.

Como em uma rede neural convencional, um mapa de Kahonen formado por vários neurônios interligados, sendo que certos neurônios recebem as entradas e propagam a informação para outros neurônios até que um neurônio de saída seja ativado. A diferença de um mapa de Kahonen é que ele é treinado usando Competitive Learning².

Competitive Learning [2] é um método de treinamento não-supervisionado onde neurônios competem pelo direito de representação um subconjunto dos dados da entrada. Ou seja, as propriedades topológicas dos dados de entrada são preservadas. Este método funciona aumentando a especialização em cada neurônio da rede, fazendo com que seja ideal para clustering.

IV. CONJUNTO DE DADOS

O conjunto de dados utilizado é a coleção de músicas Marsyas GTZAN (<https://github.com/JustGlwing/minisom>). Este conjunto foi coletado de 2000-2001 de diversas fontes e contém 1000 segmentos de 30 segundos cada. Há 10 gêneros diferentes com 100 músicas para gênero.

V. EXTRAÇÃO DAS CARACTERÍSTICAS

As características das músicas foram extraídas utilizando técnicas de processamento de sinais digitais encontradas na biblioteca Marsyas. As seguintes características foram extraídas:

- Andamento: É o grau de velocidade da música. Apesar de variar muito mesmo entre músicas do mesmo gênero, a comparação do andamento de um conjunto de dados pode diferenciar gêneros parecidos (Exemplo: Diferenciar o rock clássico do rock pesado).
- Tom: Se refere a nota em relação à qual se constrói as escalas da música. Este elemento tende a ser bom para

¹Marsyas é uma biblioteca livre para processamento de sinais digitais. <http://marsyas.info/>

²Normalmente uma rede neural é treinada utilizando métodos de correção de erros como propagação para trás ou método gradiente.

comparações pois músicas do mesmo gênero tendem a usar as mesmas escalas ou escalas parecidas (Exemplo: Escala pentatônica no gênero Blues).

- Tempo: É a pulsação básica subjacente de uma composição qualquer. Não se deve confundir tempo com andamento, onde tempo é referente ao "tempo musical" e o andamento é referente ao "tempo real".
- Timbre médio: O timbre é a qualidade de uma nota musical que difere tipos de produções de som. Está relacionado com a fonte/instrumento da música.
- Média e desvio padrão dos MFCCs (Mel-frequency Cepstral Coefficients): São coeficientes para representação do som baseado na percepção auditiva humana. Eles são derivados a partir de transformadas inversas de Fourier, aplicada nas janelas de frequências da música.

VI. IMPLEMENTAÇÃO DO MAPA DE KAHONEN

A implementação utilizada para o mapa de Kohonen foi a Minisom, que pode ser encontrada em <https://github.com/JustGlowing/minisom>. Sobre essa implementação:

“Minisom é uma implementação minimalista baseada em Numpy de um mapa de Kohonen. Um mapa de Kohonen é um tipo de rede neural capaz de converter relações complexas não-lineares entre itens em relações geométricas simples.”

VII. METODOLOGIA

Para os testes, redes de tamanho 7x7 até 15x15 foram utilizadas. Com $\sigma = 3.0$, taxa de aprendizagem 0.5 e com método de treinamento aleatório usando 100 iterações.

Os gêneros utilizados para os testes foram:

- Clássico
- Eletrônico
- Folk
- Rock Clássico
- Reggae

Para cada gênero, 100 músicas foram utilizadas. Vários testes foram feitos usando permutações diferentes das características disponíveis e gêneros musicais para fazer uma análise sobre quais características são melhores para o conjunto de dados utilizados.

VIII. RESULTADOS

A. Andamento, Tom e Tempo

Os primeiros testes feitos utilizaram as características mais teóricas das músicas- Andamento, Tom e Tempo - por terem relações fortes entre si.

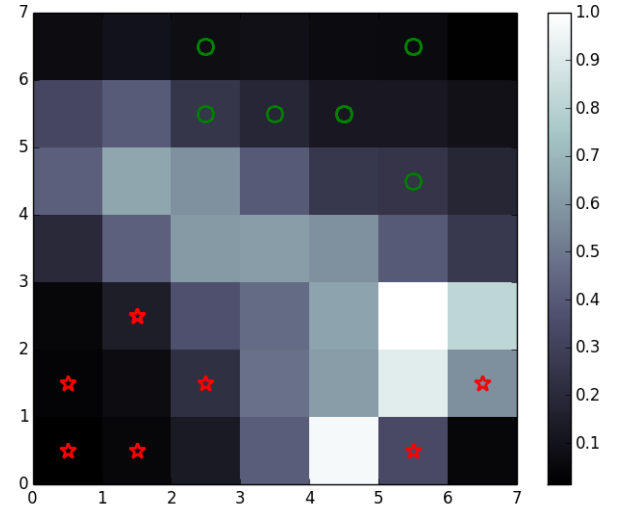


Figura 1. Rock em Vermelho, Eletrônico em Verde.

A figura 1 mostra o mapa da comparação entre os gêneros Rock e Eletrônico. É possível ver uma clara divisão entre os gêneros, apesar de haver uma certa divisão dentro do gênero Rock. Isso é mais acentuado no resultado da comparação entre Rock e Folk na figura 2.

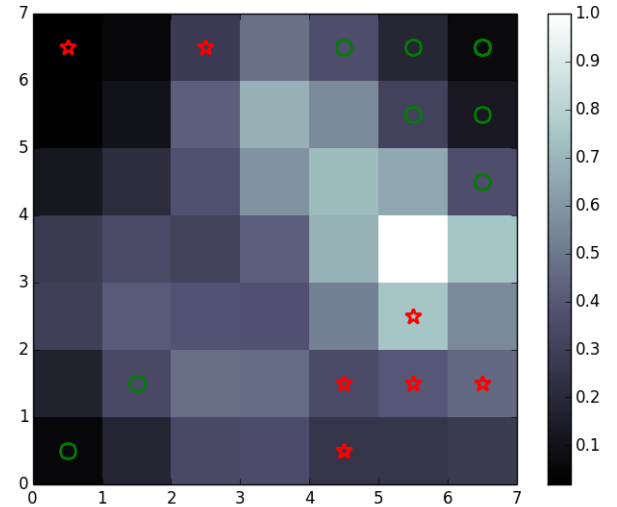


Figura 2. Rock em Vermelho, Folk em Verde.

Isso acontece quando a diferença com relação as características de cada grupo é muito pequena. Logo, qualquer pequena diferença acaba gerando um novo grupo de neurônios. É possível identificar isso nos resultados pois Rock e Folk são muito

mais parecidos que Rock e Eletrônico, logo a divisão é menos precisa.

Nas figuras 3 e 4 temos divisões mais precisas. No caso do Rock e do Reggae, as características são bem distintas entre os grupos, e caso do Clássico, Rock e Folk a diferença não é tão grande entre os dois grupos Rock e Fol), mas é com relação ao Clássico.

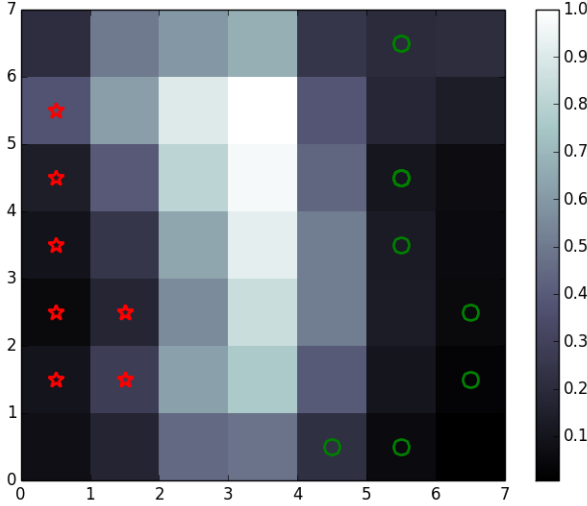


Figura 3. Rock em Vermelho, Reggae em Verde.

avelmente visíveis para essas características, há casos em que elas não funcionam bem (Figura 5). Parte das músicas do gênero Clássico estão no cluster do gênero Eletrônico, e sonoramente as músicas são muito distintas. Isso ocorre pois as características não estão representado bem este conjunto de músicas.

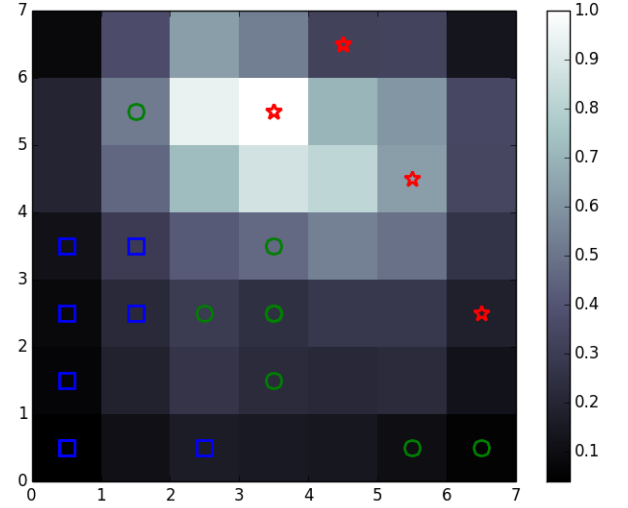


Figura 5. Folk em Vermelho, Clássico em Verde, Eletrônico em Azul.

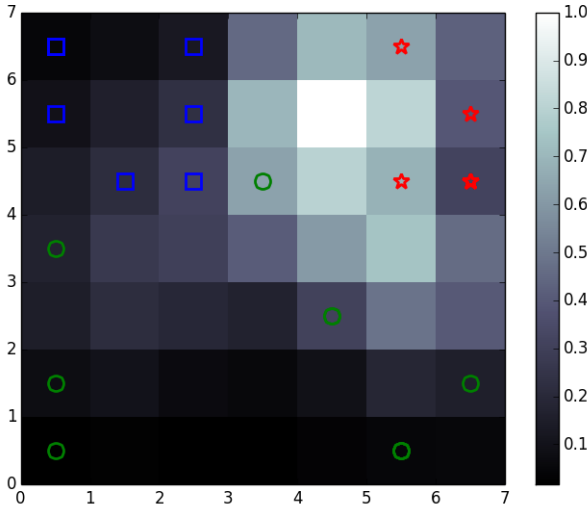


Figura 4. Clássico em Vermelho, Rock em Verde, Folk em Azul.

Apesar de nos casos anteriores as divisões serem razo-

B. Timbre

As figuras 6 e 7 mostram os resultados do mapa de comparações utilizando o timbre extraído das músicas. Os resultados para a comparação entre Rock, Clássico e Folk é parecido com quando foi utilizado o andamento, tom e e tempo. Porém na comparação entre Clássico, Folk e Eletrônico a divisão é bem imprecisa, com geração de vários pequenos grupos.

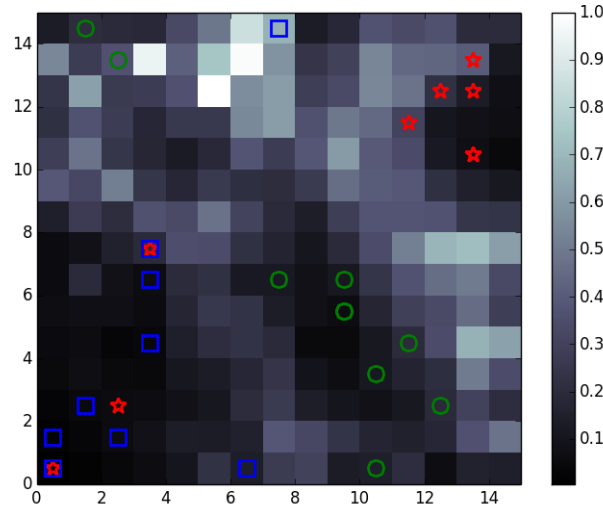


Figura 6. Rock em Vermelho, Clássico em Verde, Folk em Azul.

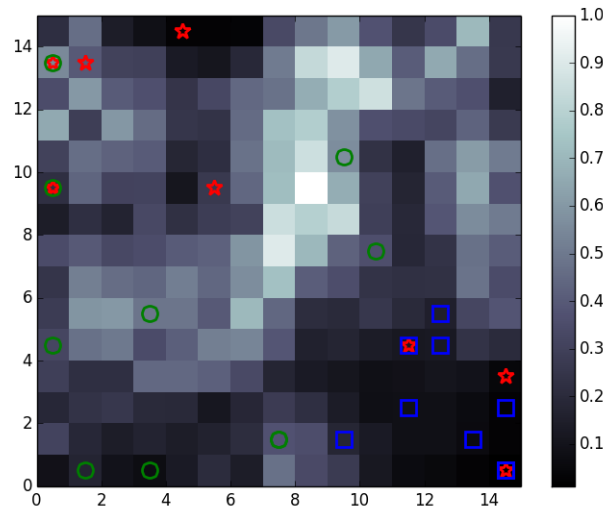


Figura 7. Clássico em Vermelho, Folk em Verde, Eletrônico em Azul.

No geral, o timbre se mostrou uma característica inconsistente, principalmente com comparações entre gêneros que utilizam instrumentos musicais mais variados. Parte do problema com o uso do timbre se deve a etapa de extração do timbre como característica, pois é uma das informações que mais se perdem durante a transformação de físico para digital (Gravação de uma música).

C. MFCCs

As comparações feitas utilizando MFCCs como características de similaridade resultam em grupo bem definidos, com diferenciações precisas entre os gêneros musicais (Figuras 9, 8 e 10). Como o MFCC se baseia na frequência, as diferenças são mais marcantes, inclusive com gêneros parecidos como Rock e Folk. É possível dizer que os MFCCs fazem uma divisão melhor pois eles não utilizam informações separadamente. Ao contrário, a frequência é afetada por todas as outras características citadas.

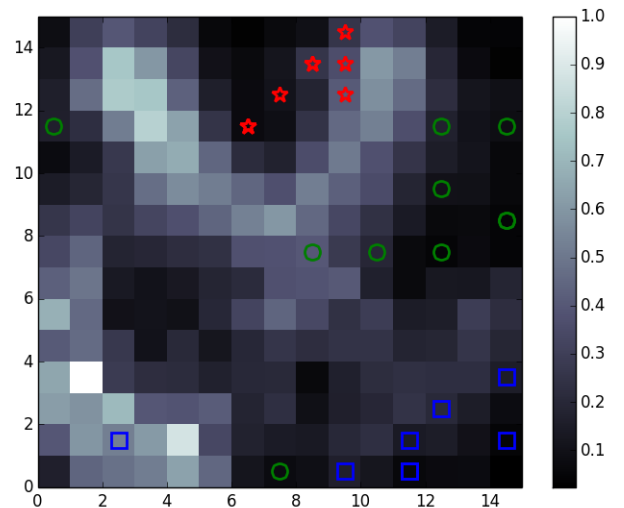


Figura 8. Rock em Vermelho, Reggae em Verde, Folk em Azul.

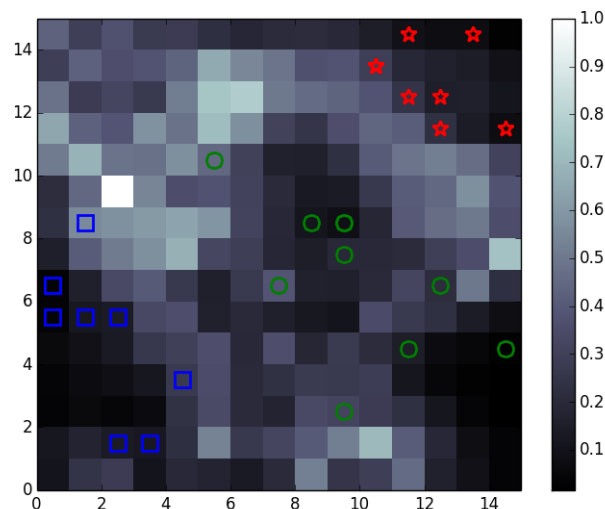


Figura 9. Rock em Vermelho, Clássico em Verde, Eletrônico em Azul.

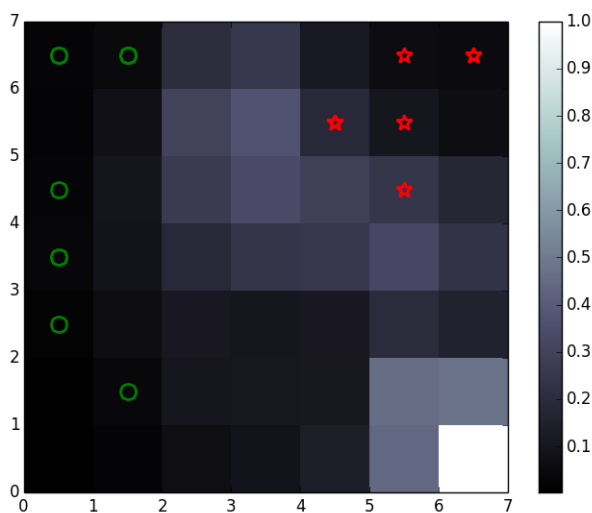


Figura 10. Clássico em Vermelho, Eletrônico em Verde.

peso para os MFCCs) tem resultados ainda mais precisos. Pelo resultado dos experimentos é possível ver que o timbre não é uma boa característica para se comparar gêneros musicais, resultando em mapas com divisão pouco precisa.

REFERÊNCIAS

- [1] Teuvo Kohonen *Self-Organized Formation of Topologically Correct Feature Maps*, Biological Cybernetics 43 (1): 59–69, 1982.
- [2] David Rumelhart, David Zipser, James L. McClelland *Parallel Distributed Processing*, Vol. 1. MIT Press. pp. 151–193, 1986.
- [3] G. Klemm, R. Valle, T. Lima, *Automatic Support for Digital Mashups*, 2014. <https://github.com/giancarlokc/Automated-Meshup/blob/master/paper.pdf>.

IX. CONCLUSÃO

O uso do mapa de Kohonen para clustering de músicas com relação a seus gêneros se provou razoavelmente eficiente, gerando grupos bem definidos na maioria dos casos.

Mais especificamente, os MFCCs se mostraram as melhores características para essa comparação de similaridade. O conjunto de características Tempo, Andamento e Tom também teve bons resultados, embora um pouco menos precisos. A utilização destes em conjunto com os MFCCs (Com maior