

Modelo automático de classificação de gêneros musicais amazônicos

Douglas Silva¹, Lucas Zampar¹, Felipe Rodrigues¹, Cláudio Gomes¹

¹Departamento de Ciências Exatas e Tecnologias - Universidade Federal do Amapá (UNIFAP)
Caixa Postal 261 – 68.906-970 – Macapá – AP – Brazil

{douglasdasilvasousa8, lucas.26.zampar, leepe.fr}@gmail.com, claudiorogério@unifap.br

Resumo. A globalização afeta a preferência musical da sociedade atual, que apresenta continua estima por ou gêneros musicais internacionais em detrimento aos nacionais ou locais. A música é dos meios de comunicação utilizados para a construção e organização da estrutura social influenciando o estilo de vida, gostos e convivência interpessoal. Portanto, cada manifestação de gênero musical apresenta função distinta para um ouvinte, como por exemplo, para dançar, festejar, descançar, ajudar na solidão, na tristeza etc. Diversos aplicativos como spotify e soundcloud, utilizam-se de classificadores de gêneros musicais para indicar, prever ou sugerir novas músicas para seus ouvintes. Por diversos motivos, a maioria dos classificadores não possui informações de gêneros musicais regionais. Este trabalho propõem um modelo de classificação automático de gêneros musicais populares amazônicos. Inicialmente, confeccionou-se uma base de dados contendo os gêneros musicais: andino, brega, carimbó, cúmbia, marabaixo, pasillo, salsa e vaqueirada, oriundos da região da Amazônia Legal dos países: Brasil, Guiana Francesa, Venezuela, Colômbia, Equador, Bolívia e Peru. Para a construção da base de dados, extraiu-se diversas características de cada música ao total de 64 parâmetros. Analisou-se os modelos de aprendizado de máquina na qual XGB, KNN, SVM e MLP obtiveram acurácia de 67.62%, 74.12%, 71.35%, 76.13%, respectivamente.

Palavras-chave — aprendizado de máquina, recuperação da informação musical, sistemas de classificação

Abstract. Globalization affects the musical preference of today's society, which has continued esteemed by international rhythms compared to the national or local. Music is a form of communication used for the construction and organization of the social structure, which It influences lifestyles, tastes, and interpersonal relationships. Therefore, each musical genre manifestation has different functions for a listener, for example, to dance, celebrate, rest, help in solitude, sadness, etc. Several applications use music genre classifiers to indicate, preview, or suggest new music for their listeners. For various reasons, most classifiers do not have information on regional music genres. This work proposes an automatic classification model for Amazonian popular musical genres. Initially, a database was created containing the musical genres: andino, brega, carimbó, cúmbia, marabaixo, pasillo, salsa e vaqueirada, from the Legal Amazon region of the countries: Brazil, French Guiana, Venezuela, Colombia, Ecuador, Bolivia and Peru. For the construction of the database, several characteristics of each song were extracted from a total of 64 parameters. Machine learning models were analyzed in which XGB, KNN, SVM e MLP obtained an accuracy of 67.62%, 74.12%, 71.35%, 76.13%, respectively.

Keywords — machine learning, music information retrieval, clas-

sifiers

1 Introdução

A globalização afeta a preferência musical da sociedade que tem continuo estima por ritmos ou gêneros musicais internacionais em detrimento à nacionais ou locais. A música é uma forma de expressão invisível, não palpável e ainda maleável tornando-se difícil a tarefa de descrevê-la, rotulá-la ou classificá-la. A base de qualquer música é composta por o ritmo, melodia e harmonia. O ritmo é a pulsação contínua que manifesta a marcação temporal regular de uma música, como por exemplo bater palmas ou estalar os dedos [1]. Além do ritmo, há ainda a melodia que é uma sequência de notas tocadas por instrumentos. Por fim, o último elemento básico da música é a harmonia, responsável em acompanhar a melodia a partir de notas tocadas simultâneas e sequenciais (acordes). A harmonia dá sentimento para a música, é responsável em gerar músicas com aspectos alegres, melancólicos, tristes etc.

O autor Menezes [2] comenta que o Brasil consome em maior quantidade, música popular brasileira que a música internacional, indicando a possibilidade de “reinvenção brasileira” a cada momento. Menezes ainda afirma a relevância da contribuição da música popular brasileira para a compreensão sobre as músicas populares do mundo. Dessa forma, o estudo sobre músicas regionais populares da Amazônia tem grande valor para o cenário musical regional, nacional e internacional.

Com base em intenções, objetivos ou valores históricos e culturais, um gênero musical é caracterizado a partir de propriedades comuns ou similares relacionadas à instrumentação, estrutura rítmica e conteúdo harmônico extraídos a partir de várias músicas [1]. Portanto, cada manifestação rítmica apresenta funções distintas para o ouvinte, como por exemplo, para dançar, festejar, descançar, ajudar em momentos de solidão, na tristeza etc.

Atualmente, diversos aplicativos utilizam-se de classificadores para indicar, prever ou sugerir novas músicas para seus ouvintes. Basicamente, a cada nova interação do usuário ao aplicativo, os classificadores adquirem novos conhecimentos e aprendem novas opções musicais para os ouvintes. Assim, tanto o usuário quanto os classificadores são duplamente consumidores de um sistema com propósitos similares e complementares [1].

Para classificar gêneros musicais é necessário, inicialmente, confeccionar uma base de dados, que em grande parte é constituído de ação manual, para prosseguir através de diversos modelos computacionais de processamento e classificação. Na literatura há várias arquiteturas que realizam o processo automático de classificação de gêneros musicais, porém a maioria não possui informações de gêneros musicais regionais ou, por diversos motivos, estão apenas engajados na classificação de gêneros musicais internacionais [2]. Assim, ouvintes que buscam ritmos regionais ou locais podem ter grande desconforto de sugestões por limitações de classificadores [3].

Este trabalho propõem um modelo de classificação de gêneros musicais amazônicos. Inicialmente, criou-se uma base de dados a partir da plataforma de *streaming youtube* contendo os ritmos musicais da amazônicos envolvendo os países: Brasil, Guiana Francesa, Venezuela, Colômbia, Bolívia e Peru. Os principais ritmos e gêneros analisados foram: marabaixo, brega, vaqueirada, carimbó, cumbia, merengue, salsa, pasillo e andino.

O artigo está organizado na seguinte forma. A seção 2 apresenta discussões de trabalhos relacionados. Na seção 3, apresenta-se o classificador automático de ritmos amazônicos e os resultados obtidos e, por fim, a seção 4, comenta-se as conclusões e trabalhos futuros.

2 Trabalhos Relacionados

O trabalho de Guggari et al. [4] apresenta uma abordagem de classificação de ritmos a partir de árvore de decisão. Realizou comparativos entre os algoritmos de floresta aleatória, C4.5, bagging e adaboost em 15 base de dados. A quantidade de classes e parâmetros utilizados para a extração de características musicais foi dependente de casa base de dados. Comparando-se as técnicas de classificação com as bases de dados, obteve-se diferentes resultados, em que no pior ao melhor caso alcançou-se 50% a 100% de acurácia. Porém, não pode-se afirmar qual a técnica mais indicada para a criação de um classificador de ritmos, pois cada base de dados possuiam quantidade de amostras, parâmetros e classes de ritmos diferentes.

O trabalho de Karunakaran e Arya [5] apresenta um classificador baseado nas técnicas de K-vizinhos mais próximos, máquina de vetor suporte, *naive Bayes*, redes neurais, análise do discriminante quadrático e lógica *fuzzy*. Investigou as classes de pop, rock e músicas eletrônicas nas bases de dados GTZAN [6] e *Free Music Archive* [7] com acurácia de 90% e 70% respectivamente.

O trabalho de Silla Jr. et al. [8] realizou um classificador automático de músicas latina. Para isso, confeccionou a base de dados intitulada *Latin Music Database (LMD)* [9] que possui 3160 músicas de 10 gêneros musicais: tango, bolero, bachata, salsa, merengue, axé, forró, sertaneja, gaúcha e pagode. Esta base possui três opções: utilização de 30 segundos iniciais, 30 segundos intermediário e 30 segundos finais de cada música. O classificador apresentou melhor resultado com características utilizando o segmento intermediário de cada da música. Realizou vários comparativos entre os modelos de redes neurais, máquina de vetor suporte e *naive bayes* na qual dependendo do cenário obteve em seu melhor caso a acurácia de 65,73%.

3 Classificador de ritmos populares da amazônia

Inicialmente, para classificar gêneros musicais é necessário, confeccionar uma base de dados, que em grande parte é substituído de ação manual, para em seguida, prosseguir a etapa de classificação por diversos modelos computacionais.

Na literatura há vários modelos que realizam o processo automático de classificação de gêneros musicais, porém a maioria não possui informações de gêneros musicais populares da amazônia ou, por diversos motivos, estão apenas engajados na classificação de gêneros musicais internacionais. Assim, ouvintes que buscam ritmos regionais ou locais podem ter grande desconforto de sugestões por limitações de classificadores.

A subseção 3.1 descreve características dos gêneros musicais escolhidos. A subseção 3.2 apresenta as características

utilizadas para a construção da base de dados dos gêneros musicais populares da amazônia. E, por fim, a subseção 3.3 apresenta as abordagens utilizadas para o modelo de classificador de gêneros musicais da amazônia.

3.1 Seleção dos gêneros musicais populares da Amazônia

Dessa forma, a partir da plataforma de *streaming youtube* realizou a seleção de, em média, 100 músicas para cada ritmo de: andino, brega, carimbo, cumbia, marabaixo, merengue, pasillo, salsa e vaqueirada. As músicas andinas são gêneros populares na Cordilheira dos Andes que inclui os países da Bolívia, Equador e Peru. Para cada gênero, utilizou-se mais de 6 grupos musicais diferentes.

Trazido de zona rural da Colômbia, a cumbia possui várias vertentes regionais em todo o país com características de folclore e celebrações festivas. Nos países da Colômbia, Panamá, Guianas, República dominicana, Panamá e região norte do Brasil o merengue apresenta conotação de celebração festiva bastante popular. O pasillo é outro ritmo de origem colombiana com grandes variações encontradas nas, principalmente, nos países da Colômbia, Panamá e Equador. A salsa é um dos ritmos mais bailados do mundo com várias influências rítmicas com rica sonoridade nos países da Colômbia, Porto Rico, México, Venezuela e Cuba.

O gênero do brega é um ritmo característico da região norte do Brasil, especificamente do estado do Pará, com influências de vários gêneros musicais com características dançantes oriundos da América Central, reconhecido como patrimônio cultural do Brasil. Outro ritmo da região norte do Brasil é o carimbó que apresenta características indígenas e influências das cumbias de folclore, é reconhecido como patrimônio cultural do Brasil. O marabaixo, reconhecido como patrimônio cultural do Brasil, é um ritmo popular da região norte do Brasil, especificamente no estado do Amapá, sobre influências rítmicas africanas e indígenas. A vaqueirada, ritmo popular do estado de Amazonas no norte do Brasil, apresenta características de ritmos de ciranda, toada e com encenação popular com diversas influências indígenas, africanas também reconhecido como patrimônio cultural do Brasil.

3.2 Criação da base de dados

A partir dos gêneros selecionados, comentados na subseção 3.1, extraiu-se características do tempograma, *chroma short-time Fourier transform* (*chroma.stft*), *chroma Constant Q spectral Transform* (*chroma.cqt*), RMS e *zero crossing rate (ZCR)* ao total de 64 parâmetros. Alguns desses parâmetros citados possuem diversos dados de um vetor ou matriz, dessa forma realizou-se testes estatísticos de média, moda e mediana criando 3 bases de dados, uma para cada teste estatístico. Ao final, a mediana apresentou melhor acurácia comparada com a média e moda.

O tempograma indica uma matriz de prevalência de certo tempo para cada momento temporal. Os testes mostraram que o uso do tempograma tiveram melhor acurácia caso fossem utilizados isoladamente. Portanto, o tempograma auxilia na identificação da velocidade dos ritmos. O *chroma.stft* apresenta características relacionadas às oitavas musicais, assim, auxilia na caracterização de tonalidades musicais. O *chroma.cqt* realiza de modo semelhante ao *chroma.stft* porém em escala logarítmica com características estatísticas para cada amostra no tempo, considerando questões de articulação, arpeggio e acordes musicais. Apesar de apresentarem características conceituais próximas, os

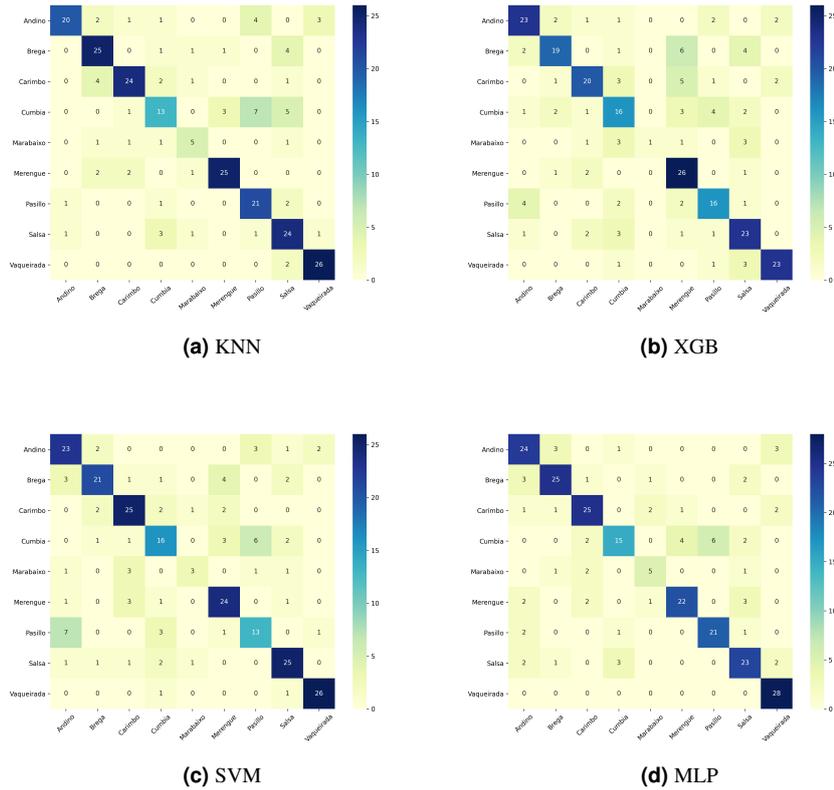


Figura 1: Matriz de confusão de testes entre modelos computacionais

Parâmetro	Índices das matrizes
Tempograma	[1 - 45]
Chroma stft	[1 - 5]
Chroma cqt	[1 - 12]
RMS	[1]
ZCR	[1]

Tabela 1: Índices utilizados da matriz de cada parâmetro

resultados alcançados em conjunto foram melhores caso utilizasse separadamente. Utilizou-se o *root-mean-square* (rms) devido a grande maioria das músicas não estarem equalizadas e por fim o *zero crossing rate* realiza avaliação de mudanças rápidas do sinal pelo cruzamento por zero. A tabela 1 apresenta os índices utilizados para cada parâmetro.

3.3 Classificador de gêneros musicais amazônicos

Conforme as considerações descritas na subseção 3.2, gerou-se a base de gêneros musicais populares amazônicos¹. Em seguida, realizou-se um estudo dos parâmetros relevantes para o classificador, conforme a subseção 3.2. Em sequência, realizou-se as análises de modelos computacionais entre *KNN*, *gradiente boosting*, *multilayer perceptron* e máquina de vetor suporte (*SVM*). Para todos os modelos, utilizou-se 70% da base de dados como treino e 30% como testes.

¹<https://bit.ly/3keCFuK>

No modelo *KNN* a partir de testes iniciais, obteve-se melhores resultados com os hiperparâmetros de 10 *k-vizinhos* e com análise de parâmetros considerando *distância*. O *gradiente boosting* (*XGB*) é um modelo computacional que realiza pequenas previsões para um conjunto de dados corrigidos temporalmente o preditor final. Assim, o *XGB* analisa a contribuição de cada preditor sequencial através de seu crescimento baseado no erro residual do preditor anterior. Com relação modelo de máquina vetor suporte (*SVM*), tenta encontrar a distância de linhas para a separação entre os dados das classes. O *SVM* é indicado para um volume de dados não muito grande, que não estejam sobrepostos, ou ainda com pouca quantidade de ruídos. Para o *SVM*, utilizou-se a função de decisão *ovr*, *kernel rbf* com taxa de 0.00205 e com penalidade *L2*.

Com relação à redes neurais *perceptron* (*MLP*), é outro modelo computacional de decisão com objetivos de classificar, agrupar e ainda prever informações a partir de dados repassados a rede de neurônios que com suas interconexões realizam o aprendizado do dados. É um modelo bastante complexo e inerente de sua estrutura possuindo diversas variáveis controláveis para facilitar resultados esperados [10]. Para o classificador, utilizou-se uma rede com 1520, 300, 100 por 9 camadas escondidas utilizando *tangente hiperbólica*, *relu*, *tangente hiperbólica* e *softmax* com 100 épocas e *batch size* de tamanho 32.

A figura 1 apresenta os resultados de cada modelo computacional. O *XGB*, *KNN*, *SVM* e *MLP* obtiveram acurácia de 67.62%, 74.12%, 71.35%, 76.13%, respectivamente. Em todos os modelos, percebe-se a dificuldade em classificar gêneros musicais do marabaixo, pois teve-se apenas o total de 30 músicas. Apresentou-se dificuldade em encontrar músicas do marabaixo. Com relação à cumbia e merengue, percebe-se algumas falhas na

classificação, pois esses gêneros influenciam os demais gêneros selecionados.

Dessa forma, os modelos KNN e SVM que possuem por essência realizar separação de classes apresentaram resultados semelhantes com relação ao resultado de acurácia quanto aos erros de classificação. Já os modelos XGB e MLP com características de aprendizado temporal, apresentaram evoluções de classificação dependendo de seus respectivos hiperparâmetros. No geral, analisando os erros dos classificadores pode-se considerar aceitáveis pois, os ritmos que foi classificados possuem sonoridade próxima umas das outras como a brega, carimbó, cúmbia, merengue e pasillo.

4 Conclusões e trabalhos futuros

Conclui-se que o modelo de redes neurais apresentou melhores resultados com relação aos demais modelos. Para trabalhos futuros, pretende-se analisar outros modelos de aprendizado de máquina com objetivo de alcançar melhorias na acurácia da classificação e ainda desenvolver melhor compreensão entre os ritmos populares da amazônia selecionados. Além disso, adicionar novas músicas para cada gênero musical afim de melhor compreender os ritmos por diversas características computacionais.

Referências

- [1] Susan Schreibman, Ray Siemens, and John Unsworth. *A new companion to digital humanities*. John Wiley & Sons, 2015.
- [2] Rafael José de Menezes Bastos. *As contribuições da música popular brasileira às músicas populares do mundo: diálogos transatlânticos Brasil-Europa-África - Segunda parte*. Universidade Federal de Santa Catarina, 2007.
- [3] Andre Holzapfel, Bob Sturm, and Mark Coeckelbergh. Ethical dimensions of music information retrieval technology. *Transactions of the International Society for Music Information Retrieval*, 1(1):44–55, 2018.
- [4] Shankru Guggari, Vijayakumar Kadappa, V Umadevi, and Ajith Abraham. Music rhythm tree based partitioning approach to decision tree classifier. *Journal of King Saud University-Computer and Information Sciences*, 2020.
- [5] Nagamanoj Karunakaran and Arti Arya. A scalable hybrid classifier for music genre classification using machine learning concepts and spark. In *2018 International Conference on Intelligent Autonomous Systems (ICoIAS)*, pages 128–135. IEEE, 2018.
- [6] Bob L. Sturm. The GTZAN dataset: Its contents, its faults, their effects on evaluation, and its future use. *CoRR*, abs/1306.1461, 2013.
- [7] Michaël Defferrard, Kirell Benzi, Pierre Vandergheynst, and Xavier Bresson. Fma: A dataset for music analysis. *arXiv preprint arXiv:1612.01840*, 2016.
- [8] Carlos N Silla Jr, Celso AA Kaestner, and Alessandro L Koerich. Automatic music genre classification using ensemble of classifiers. In *2007 IEEE International Conference on Systems, Man and Cybernetics*, pages 1687–1692. IEEE, 2007.
- [9] Carlos Nascimento Silla Jr, Alessandro L Koerich, and Celso AA Kaestner. The latin music database. In *ISMIR*, pages 451–456, 2008.
- [10] Aurélien Géron. *Mãos à Obra: Aprendizado de Máquina com Scikit-Learn & TensorFlow*. Alta Books, 2019.