

Estudo da Relevância das Características num Sistema Neural de Classificação do Gênero Musical

Juliana Lopes, Yure Souza, Eduardo Simas Filho, Claudia Farias.

Grupo de Pesquisa em Ensaios Não Destrutivos, Instituto Federal da Bahia.
{jlopes, yuresouza, esimas, cfarias}@ifba.edu.br

Resumo – Considerando o problema do acesso automático ao conteúdo de arquivos de áudio, este trabalho realiza um estudo da relevância das características estimadas dos sinais para a eficiência de um classificador de gêneros musicais. Foi utilizado um discriminador baseado em Redes Neurais Artificiais e foram estimadas, ao todo, 28 características de cada arquivo da base de dados, que era composta por 837 arquivos pertencentes a 10 ritmos musicais diferentes. Após o projeto do classificador utilizando todos os parâmetros disponíveis, foi iniciado um processo gradual de retirada dos atributos de entrada e re-treinamento da rede neural. Através da comparação entre os resultados, foi possível verificar quais características são mais importantes para a eficiência do sistema neural de classificação, visando a minimizar o esforço computacional do discriminador proposto.

Palavras-Chave – Redes Neurais Artificiais, Processamento Digital de Sinais, Extração de características.

1 Introdução

Nos últimos anos, juntamente com o número de usuários da *internet*, cresce também a quantidade de informações e arquivos disponíveis na rede mundial (*world wide web - www*), sendo que, boa parte destes arquivos é de áudio. Ao mesmo tempo, é cada vez maior o número de pessoas que possuem grandes coleções de arquivos de áudio, seja em computadores ou em dispositivos portáteis de reprodução (como telefones celulares e tocadores de mídia). Deste modo, sem uma ferramenta de busca e organização adequada, a localização de um conteúdo específico pode se tornar lenta e tediosa para o usuário.

Uma forma natural de classificar o conteúdo de arquivos de áudio é através do gênero musical predominante. Assim, visando a organizar efetivamente a grande quantidade de dados disponíveis e facilitar a busca do usuário por arquivos de seu interesse, cada vez mais são desenvolvidas pesquisas na área de classificação automática de gêneros musicais [1-3]. A associação de uma música a um gênero, mesmo por especialistas no assunto, nem sempre é uma tarefa fácil e exata, podendo envolver certo grau de subjetividade, o que torna a área de pesquisa muito ativa.

Os classificadores baseados em redes neurais artificiais (RNA) [4] são utilizados com sucesso em problemas complexos e de diferente natureza como física de partículas [5], medicina [6] e sistemas de potência [7]. Sistemas de classificação de arquivos de áudio utilizando redes neurais foram propostos anteriormente em trabalhos como [8, 9]. Em geral estes classificadores são alimentados por atributos estimados dos sinais de áudio numa etapa de pré-processamento (extração de características). Em geral são utilizados algoritmos de extração de características comuns ao processamento de áudio como os coeficientes *cepstrais* e a frequência fundamental, e também outros específicos para arquivos de música como o histograma rítmico [1, 2].

Este trabalho apresenta uma ferramenta de classificação musical automática que inclui 10 gêneros distintos, entre eles ritmos brasileiros como Forró e MPB (Música Popular Brasileira). É selecionado um conjunto de características considerando, na literatura existente, os atributos mais utilizados para caracterizar arquivos de áudio. Estas características são estimadas a partir de cada arquivo de música e utilizadas para treinar um classificador neural na arquitetura *perceptron* de múltiplas camadas (MLP – *Multi-Layer Perceptron*). Após o treinamento do classificador é realizado um estudo para estimar a importância de cada um dos atributos na classificação do gênero musical visando à redução do esforço computacional do sistema proposto.

O trabalho está dividido conforme descrito a seguir. Na Seção 2, é apresentado o sistema proposto para a classificação dos gêneros musicais, incluindo as rotinas de extração de características e o classificador neural utilizado. A seguir, na Seção 3 são apresentados os procedimentos utilizados para estimar a relevância das características utilizadas. Os resultados obtidos a partir da base de dados utilizada são mostrados na Seção 4. Finalmente, a Seção 5 traz as conclusões e as perspectivas para futuros trabalhos.

2 O Sistema Proposto

Neste trabalho é proposto um sistema de classificação automática de gêneros musicais que consiste em duas etapas: extração de parâmetros e classificação (através de redes neurais artificiais). Os parâmetros estimados dos sinais de áudio devem retratar as características que possibilitam a classificação do sinal em seu respectivo gênero, conforme ilustrado na Figura 1. As características utilizadas aqui serão descritas a seguir.



Figura 1: Representação esquemática do processo de classificação.

2.1. Extração de Parâmetros

Os sinais de áudio apresentam propriedades estatísticas variantes no tempo (são não-estacionários), neste caso, a extração de características deve ser realizada em janelas temporais de curta duração (garantindo a estacionariedade dentro das janelas). Na realização deste trabalho, conforme proposto anteriormente para problemas semelhantes [1-3,10], foram utilizadas janelas de *Hamming* [11] de duração igual a 30 ms com 30% de sobreposição entre janelas adjacentes. Apenas num dos conjuntos de características, o histograma rítmico, é necessário considerar toda a duração do sinal de áudio. Além do histograma rítmico foram estimados os seguintes atributos: taxa de cruzamento por zero, coeficientes mel-cepstrais, energia por faixa de frequências, centróide espectral, largura de banda e sonoridade. Alguns destes atributos foram utilizados com sucesso em um trabalho anterior [10], onde um classificador neural foi projetado para um sistema de classificação neural do gênero musical com reduzido número de classes. A seguir serão apresentados, de modo resumido, os procedimentos utilizados aqui para estimação das características mencionadas.

2.1.1. Taxa de Cruzamento por Zeros

A taxa de cruzamento zero, ou ZCR (*Zero Crossing Rate*) é comumente utilizada na caracterização de sinais de áudio [1]. Tal parâmetro é calculado através da contagem do número de vezes que o sinal temporal cruza o eixo zero [2], podendo ser visto como um descritor da frequência dominante da música (frequência de *pitch*), sendo útil também para encontrar quadros em silêncio.

2.1.2. Coeficientes Mel-Cepstrais (MFCC – *Mel Frequency Cepstral Components*)

Medidas Cepstrais fornecem informação a respeito do formato do espectro. Os coeficientes cepstrais são calculados através da transformada de Fourier inversa do logaritmo do espectro de potência do sinal [3]. Conforme indicado na Figura 2, a extração deste parâmetro é feita, inicialmente, dividindo-se o sinal em quadros através do janelamento, a seguir, a transformada discreta de Fourier (DFT – *Discrete Fourier Transform*) é aplicada em cada janela. O logaritmo do espectro de amplitude é calculado e mapeado para uma escala não-linear, a escala mel, que tenta modelar a percepção do ouvido humano às diferentes frequências de áudio.

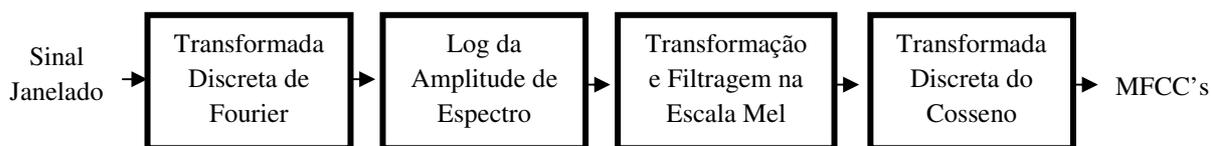


Figura 2: Fluxograma de extração dos MFCC's.

A escala mel pode ser mapeada a partir da linear (em Hz) através da Equação 1:

$$mel(f) = 1227 \times \ln\left(1 + \frac{f(Hz)}{700}\right) \quad (1)$$

Finalizando o cálculo dos coeficientes mel-cepstrais, após o mapeamento na escala mel é realizado um processo de filtragem seguido da aplicação da transformada discreta do cosseno [11].

2.1.3. Histograma rítmico

O histograma rítmico (*Beat Histogram*) visa representar o conteúdo rítmico existente num arquivo de áudio, sendo muito utilizado na classificação de gêneros musicais [12]. Esta rotina de extração de características busca encontrar a batida (ritmo) principal da música e seu período em BPM (batidas por minuto). A presença de batidas secundárias relevantes indica que a música tem um conteúdo rítmico mais intenso.

A Figura 3 descreve o procedimento para a obtenção do histograma rítmico. Inicialmente, ao sinal temporal é aplicada a Transformada *Wavelet* Discreta (DWT – *Discrete Wavelet Transform*) [11], para proporcionar uma representação compacta do sinal em tempo e frequência. Na DWT, o sinal é decomposto, a partir de filtros passa-baixa e passa-alta, em coeficientes de aproximação e de detalhe [13].

Os coeficientes da decomposição através da DWT são processados para a extração da envoltória. O processo de extração de envoltória envolve retificação de onda completa, filtragem passa-baixas, sub-amostragem e extração da média. As envoltórias de cada banda são então somadas e utilizadas para calcular a autocorrelação. Os picos da função de autocorrelação correspondem às periodicidades diferentes da envoltória do sinal.

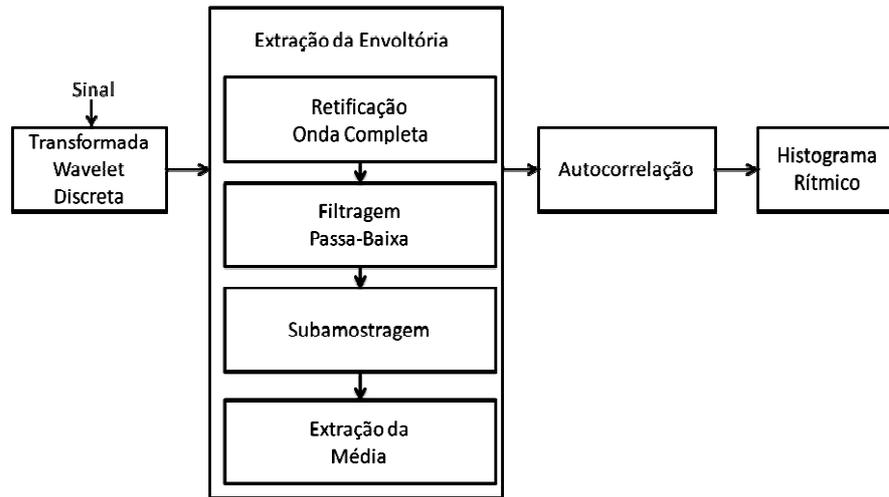


Figura 3: Fluxograma de criação do Histograma rítmico.

2.1.4. Concentração da Energia em Faixas de Frequências

Esse parâmetro foi proposto em um trabalho anterior [10] e avalia a variação da concentração das maiores amplitudes do sinal no domínio da frequência, em três bandas diferentes. Neste trabalho o espectro foi dividido em frequências baixas (0 a 600 Hz), médias (600 a 2.400 Hz) e altas (2.400 a 22.050 Hz). A energia em cada faixa é obtida a partir do seguinte procedimento (esquemático na figura 4): Inicialmente, o sinal é dividido em janelas de 30 ms, em seguida, é filtrado usando filtros FIR (*Finite Impulse Response*) [11] passa-baixa (0-600Hz) passa-faixa (600-2.400 Hz) e passa-alta (2.400-22.050 Hz). Por meio da aplicação de algoritmos de obtenção da PSD (*Power Spectral Density*), a potência do sinal é extraída em cada faixa de frequência. Então, a potência em cada banda é somada e normalizada pela energia total.

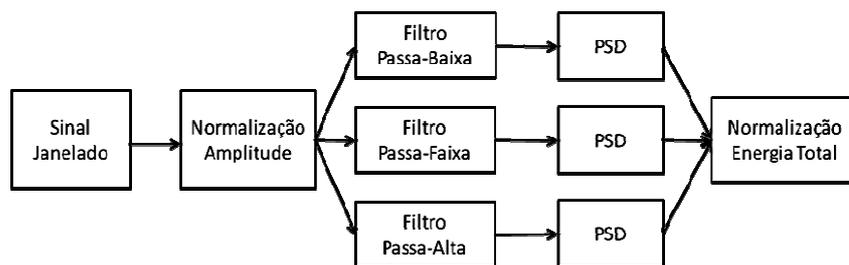


Figura 4: Etapas envolvidas no cálculo da energia concentrada em faixas de frequência.

2.1.5. Centróide Espectral

O centróide espectral representa o centro de massa do espectro de energia de cada quadro [3] e está relacionado com a percepção humana do “brilho” do sinal [8], sendo calculado a partir de:

$$ce(i) = \frac{\sum_{k=1}^k k |X_i(k)|^2}{\sum_{k=1}^k |X_i(k)|^2} \quad (2)$$

onde $X_i(k)$ são as componentes da Transformada Discreta de Fourier (TDF) do quadro i .

2.1.6. Largura de Banda

É a medida da largura da faixa de frequências presente no sinal (em Hertz).

2.1.7. Sonoridade

A sonoridade (*Loudness*) tenta reproduzir a percepção humana em relação a um sinal de áudio. O primeiro passo para o seu cálculo é modelar a resposta em frequência dos ouvidos médio e externo [3]. Uma das formas de se obter esse modelo é através da equação 3 [9]:

$$W(k) = -2,184 \cdot f(k)^{-0,8} - 6,5 \cdot e^{-0,6(f(k)-3,3)^2} + 10^{-3} \cdot f(k) \cdot 3,6 \quad (3)$$

onde a frequência $f(k)$ é medida em kHz, dada por $f(k) = k \cdot d$, e d é diferença entre duas linhas espectrais consecutivas em kHz. $W(k)$ é uma função de ponderação que atua no sinal de acordo com o comportamento auditivo, atenuando ou enfatizando as suas componentes espectrais. Assim, a sonoridade de um quadro i é calculada utilizando:

$$ld_i = \sum_{k=1}^k |X_i(k)|^2 \cdot 10^{w(k)/20} \quad (4)$$

2.1.8. O Vetor de Características

Para cada sinal de áudio o vetor de características utilizado é composto por 28 atributos conforme descrito a seguir:

- Média e variância (calculadas para todas as janelas temporais) dos parâmetros: taxa de cruzamento por zero, energia em três faixas de frequência, centróide espectral, sonoridade, largura de banda e os cinco primeiros coeficientes Mel-Cepstrais (totalizando 24 características);
- Quatro parâmetros do histograma rítmico (amplitude relativa do primeiro e do segundo picos do histograma, razão da amplitude do segundo pico dividido pela amplitude do primeiro pico, Soma do histograma (4 características)).

2.2. O Classificador Neural Utilizado

Neste trabalho foi utilizado um classificador neural na topologia *perceptron* de múltiplas camadas (MLP – *Multi-Layer Perceptron*) [4] com uma camada oculta e N neurônios na camada de saída, sendo N o número de classes do problema.

Para a determinação do número ótimo de neurônios da camada oculta, primeiramente a rede foi montada e treinada com apenas 1 neurônio oculto. Após o treinamento, foi calculada a eficiência de discriminação. A seguir, adiciona-se um neurônio oculto e os processos de treinamento e cálculo da eficiência são repetidos. O procedimento continua até que a eficiência não aumente com a adição de novos neurônios.

A eficiência global (considerando todas as classes do problema) é estimada a partir da média geométrica das eficiências individuais de cada uma das classes, conforme a expressão:

$$EF_{MG} = \sqrt{\prod_{i=1}^N EF_i} \quad (5)$$

Visando a melhor explorar durante o processo de treinamento toda a informação disponível foi realizado um processo de validação cruzada. Os exemplos disponíveis foram divididos nos conjuntos de treino, validação e teste e o treinamento da rede iniciado. A seguir uma nova divisão nos três conjuntos foi efetuada de modo aleatório e o treinamento re-iniciado. O processo continua até que 10 treinamentos com conjuntos diferentes sejam realizados.

3 Análise da Relevância das Características

O sistema proposto pode ser aplicado em dispositivos móveis de execução de áudio, como os tocadores multimídia e celulares multifuncionais. Estes dispositivos, em geral, apresentam limitado poder computacional e, deste modo é interessante verificar se é possível reduzir o custo computacional necessário para a classificação em gêneros musicais mantendo uma eficiência de discriminação equivalente.

Neste sentido, foi realizado um estudo da relevância das características utilizadas para a discriminação dos gêneros musicais. O procedimento adotado consiste na retirada de atributos individuais (ou conjuntos de atributos) e re-treino do classificador. Deste modo, pode-se verificar como a ausência de certas entradas influencia na eficiência global. Se a eficiência se mantém aproximadamente constante ou aumenta indica que a característica retirada não é relevante para o sistema. Por outro lado, a queda acentuada na eficiência de discriminação com a remoção de uma ou mais características mostra que aqueles atributos são importantes para a discriminação entre as classes.

4 Resultados

A base de dados utilizada neste trabalho é composta por 837 arquivos pertencentes a 10 ritmos musicais diferentes. A distribuição de exemplos por ritmo está disposta na Tabela 01. Os arquivos disponíveis foram divididos em conjuntos de treino, validação e teste seguindo uma proporção aproximada de respectivamente 50%, 25% e 25% do total de sinais.

Tabela 1 – Numero de amostras por ritmo

Classe	Nº de Arquivos
<i>Blues</i>	81
<i>Clássica</i>	81
<i>Forró</i>	84
<i>Hip Hop</i>	77
<i>Jazz</i>	65
<i>MPB</i>	103
<i>Pop Rock</i>	90
<i>Punk</i>	102
<i>Reggae</i>	84
<i>Soul</i>	70

Num primeiro momento, foi necessário determinar a topologia ótima para a rede neural (i.e. número de neurônios da camada oculta). Para isso, o procedimento descrito anteriormente na seção 2.2 foi adotado, variando-se o número de neurônios de 1 a 49. Pode-se observar na Figura 5 que o valor ótimo do índice de eficiência utilizado neste trabalho (média geométrica das eficiências \rightarrow EF_{MG}) foi obtido para 33 neurônios na camada oculta ($EF_{MG} = 0,6211$), embora valores semelhantes tenham sido alcançados para 9 ($EF_{MG} = 0,5834$), 25 ($EF_{MG} = 0,5999$), 31 ($EF_{MG} = 0,5938$) e 45 ($EF_{MG} = 0,6012$) neurônios.

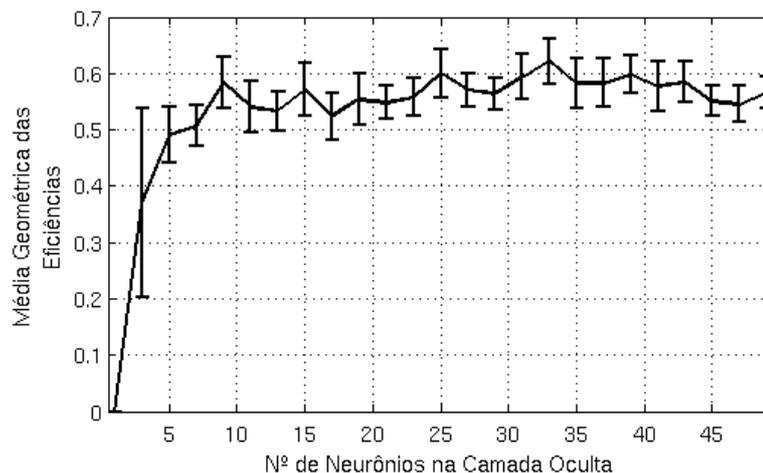


Figura 5: Variação da média geométrica das eficiências com o número de neurônios utilizado para o treinamento.

A Tabela 2 mostra a matriz de confusão obtida para o melhor caso (33 neurônios na camada oculta). Pode-se observar na coluna j da linha i a parcela da classe i associada pelo classificador à classe j . Deste modo, percebe-se que a diagonal da matriz de confusão traz as eficiências de identificação de cada uma das classes. É interessante notar que os gêneros musicais muitas vezes apresentam características semelhantes e até mesmo comuns, sendo a classificação difícil mesmo para especialistas humanos. Considerando um sistema de classificação automático, as fronteiras de decisão entre as classes se tornam sobrepostas, gerando naturalmente certo grau de confusão entre as diversas classes.

Numa análise mais detalhada da Tabela 2 é possível observar que para gêneros musicais como Jazz, MPB e Soul a eficiência de discriminação é baixa. Isso pode ser justificado considerando que estes gêneros apresentam características comuns a diversos outros, por exemplo, ~11% dos sinais de Jazz foram classificados como Blues e ~17% como MPB. Por sua vez, as músicas da classe MPB foram confundidas com as das classes Jazz (~16%), Blues (~8%), Forró (~8%) e Reggae (~8%), o que é justificado pela influência comum entre os ritmos, sendo que muitos artistas transitam facilmente entre eles. Considerando o Hip-Hop, que é um gênero caracterizado por sua batida forte e ritmada, a confusão maior aconteceu com Reggae (~18%) e Forró (~7%), que, de certo modo, também são gêneros com fortes características rítmicas. Por outro lado, os exemplos de música clássica somente foram confundidos com os de Jazz (~14% de confusão), isso pode ser explicado pela pouca influência direta da música clássica em gêneros mais populares.

Com o objetivo de verificar a relevância das características estimadas foi adotado o processo de retirada de atributos de entrada conforme descrito anteriormente na seção 3. Na Figura 6 são mostrados os resultados obtidos a partir do re-treinamento do classificador neural (mantendo agora o número de neurônios na camada oculta fixo em 33) a partir da retirada de cada uma das características utilizadas. Pode-se observar que para a retirada de duas características (C01 → Valor médio do primeiro coeficiente mel-cepstral e C20 → amplitude do segundo pico do histograma rítmico) a eficiência aumenta, indicando que elas estão apenas confundindo o classificador, não agregando informação relevante para o problema. O classificador foi re-treinado agora retirando-se ao mesmo tempo as características C01 e C20 e foi obtido $EF_{MG} = 0,6006$ (o que corresponde a ~97% do valor obtido para todas as características).

Numa análise mais detalhada, é possível verificar que ao serem retiradas individualmente as características C01, C04, C20, C21, C23 e C25 (grupo 1) a redução na eficiência é menor que 3% e retirando-se individualmente C01, C04, C14, C20, C21, C23, C25 e C27 (grupo 2) a redução na eficiência é menor que 5%. Assim, o sistema neural de classificação foi re-treinado retirando-se todas as características dos grupos 1 e 2, obtendo-se respectivamente $EF_{MG} = 0,5999$ e $EF_{MG} = 0,5970$ (correspondendo a ~96% da eficiência obtida com o uso de todas as características). Ou seja, é possível retirar 8 dos 28 atributos utilizados inicialmente para a rede neural (redução de ~30% no número de entradas) e manter ainda assim uma eficiência compatível com a obtida a partir das 28 características.

Ainda buscando avaliar a relevância das entradas, foram retirados em conjunto os atributos que são calculados a partir da mesma rotina (Ex. todos os coeficientes mel-cepstrais ou todos os parâmetros do histograma rítmico). Deste modo, o esforço computacional é reduzido de modo mais acentuado. Conforme observado na Figura 7, os conjuntos de características menos relevantes são a Largura de Banda e a Sonoridade. Retirando-se o primeiro grupo a eficiência cai para $EF_{MG} = 0,5992$ e para segundo $EF_{MG} = 0,5794$ (representando respectivamente ~96% e 93% da eficiência obtida a partir de todas as características). Pode-se perceber ainda que os conjuntos de atributos mais relevantes são (em ordem decrescente de relevância) energia em faixas de frequência, coeficientes MFCC e o histograma rítmico.

Tabela 2: Matriz de confusão obtida para o sistema neural de classificação do gênero musical.

	Blues	Clássica	Forró	Hip Hop	Jazz	MPB	POP	Punk	Reggae	Soul
Blues	67,9	0,0	3,6	0,0	7,1	10,7	7,1	3,6	0,0	0,0
Clássica	0,0	85,7	0,0	0,0	14,3	0,0	0,0	0,0	0,0	0,0
Forró	3,6	0,0	57,1	0,0	3,6	17,9	7,1	3,6	0,0	7,1
Hip Hop	0,0	0,0	7,4	66,7	0,0	0,0	3,7	0,0	18,5	3,7
Jazz	11,1	5,6	16,7	0,0	38,9	16,7	0,0	5,6	5,6	0,0
MPB	8,1	5,4	8,1	2,7	16,2	45,9	2,7	0,0	8,1	2,7
Pop	5,7	0,0	5,7	2,9	2,9	2,9	80,0	0,0	0,0	0,0
Punk	0,0	0,0	0,0	0,0	0,0	2,9	2,9	82,4	5,9	5,9
Reggae	7,1	0,0	7,1	0,0	0,0	0,0	7,1	0,0	78,6	0,0
Soul	4,2	4,2	4,2	0,0	4,2	12,5	12,5	0,0	16,7	41,7

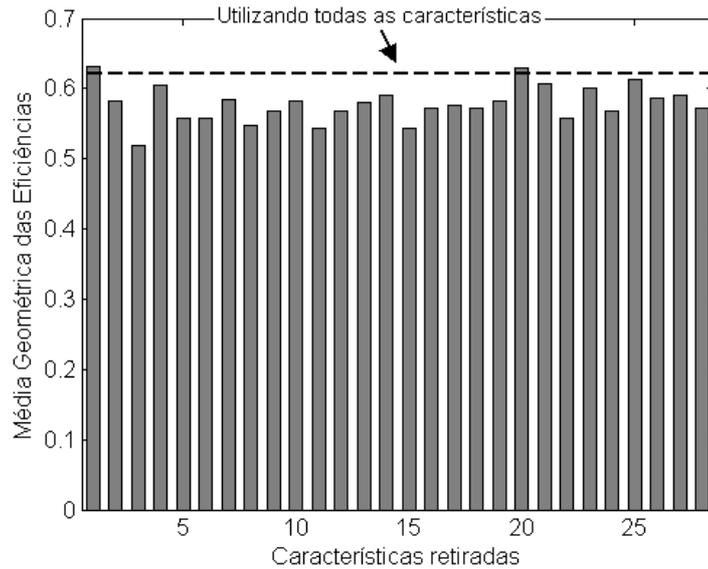


Figura 6: Variação da média geométrica das eficiências com a retirada de atributos de entrada para a rede neural.

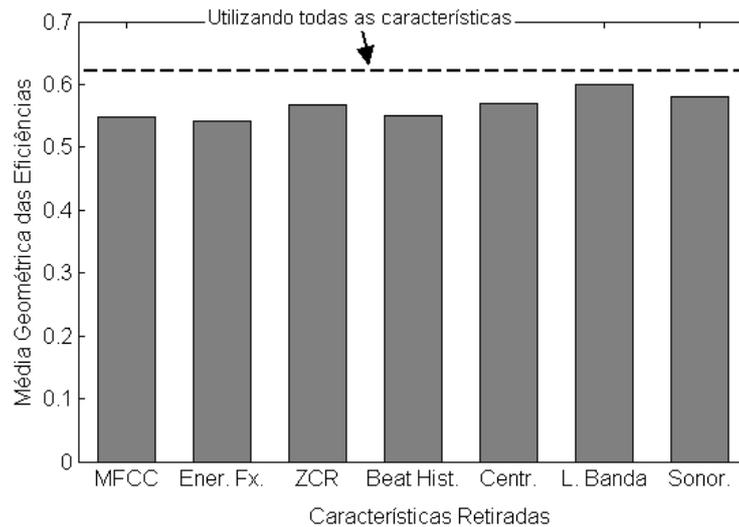


Figura 7: Variação da média geométrica das eficiências com a retirada de classes de atributos de entrada para a rede neural.

5 Conclusões

A quantidade de arquivos de áudio disponível na internet ou em coleções pessoais vem crescendo rapidamente nos últimos anos. Deste modo, são necessárias ferramentas de auxílio ao usuário na busca pelo conteúdo desejado. Este trabalho propôs um sistema automático de classificação de arquivos de áudio de acordo com o gênero musical predominante. O sistema proposto combina uma eficiente extração de características, através da estimação de atributos relevantes, com o uso de classificadores neurais através de redes tipo *perceptron* de múltiplas camadas. Com o sistema proposto foi possível obter eficiências da ordem de 80% para algumas classes e uma eficiência global (avaliada pela média geométrica das eficiências individuais) de aproximadamente 63%. A relevância de cada uma das características estimadas foi avaliada a partir de um processo de re-treinamento da rede com a gradual retirada de atributos de entrada. Assim, foi verificado que é possível otimizar o custo computacional do sistema proposto a partir da redução do número de atributos de entrada em aproximadamente 30% e ainda obter uma eficiência média da ordem de 60%. Em trabalhos futuros pretende-se avaliar o grau de dependência estatística entre os atributos utilizados e, caso seja necessário, utilizar uma transformação como a análise de componentes independentes (ICA – *Independent Component Analysis*) [14] numa etapa de pré-processamento à classificação, visando eliminar a dependência estatística mútua entre as entradas da rede neural, o que pode contribuir para facilitar o processo de treinamento. Outra linha de trabalho é a adoção de regras de decisão *fuzzy* no classificador neural visando a melhor considerar as semelhanças existentes entre os diversos gêneros musicais que acabam gerando sobreposição nas fronteiras de decisão entre classes. É importante

ainda realizar um estudo quantitativo a respeito da redução do custo computacional com a retirada de cada uma das características.

Agradecimentos

Agradecemos ao IFBA pela infraestrutura, à FAPESB pelo apoio financeiro e ao GPEND, pelo apoio técnico.

Referências

- [1] Lu, L., Zhang, H. J. and Jiang, H, Content Analysis for Audio Classification and Segmentation, *IEEE Transactions on Speech and Audio Processing*, (2002), Vol. 10, No. 7.
- [2] Kim, H.-G., Moreau, N. and Sikora, T (2005), *MPEG-7 Audio and Beyond: Audio Content Indexing and Retrieval*, John Wiley & Sons, Ltd.
- [3] Pereira, E. M. (2009), *Estudos sobre uma Ferramenta de Classificação Musical*, dissertação de mestrado em Engenharia Elétrica, UNICAMP, Campinas-SP.
- [4] S. Haykin. *Neural Networks and Learning Machines*. Prentice Hall, New Jersey, United States, third edition, 2008.
- [5] A. dos Anjos, R. Torres and J. Seixas. “Neural triggering system operating on high resolution calorimetry information”. *Nuclear Instruments and Methods in Physics Research A*, vol. 559, no. 1, pp. 134–138, 2006.
- [6] R. Silipo and C. Marchesi. “Artificial neural networks for automatic ECG analysis”. *IEEE Transactions on Signal Processing*, vol. 46, no. 5, pp. 1417–1425, 1998.
- [7] Z.-L. Gaing. “Wavelet-based neural network for power disturbance recognition and classification”. *IEEE Transactions on Power Delivery*, vol. 19, no. 4, pp. 1560–1568, 2004.
- [8] Qian Ding and Nian Zhang, Classification of Recorded Musical Instruments Sounds Based on Neural Networks, , *Proceedings of the 2007 IEEE Symposium on Computational Intelligence in Image and Signal Processing (CIISP 2007)*.
- [9] Thiede, T. V. (1999) *Perceptual Audio Quality Assessment using a Non-Linear Filter Bank*, PhD thesis, Technical University of Berlin, Berlin.
- [10] Borges, E., Simas Filho, E., Farias, C., Ribeiro, I., Lopes, D., Classificação do Gênero Musical Utilizando Redes Neurais Artificiais, *X Congresso Norte-Nordeste de Pesquisa e Inovação*, (2010).
- [11] Diniz, P. S. R., da Silva, E. A. B. e Lima Netto, S. (2004), *Processamento Digital de Sinais*, Ed. Bookman, Porto Alegre.
- [12] Tzanetakis, G. and Cook, P. (2002), Musical genre classification of audio signals, *IEEE Transactions on Speech and Audio Processing*, 10(5):293-302.
- [13] Meng A. (2006), *Temporal feature integration for music organization*, Ph.D, dissertation, Informatics and Mathematical Modeling Technical Univ. Denmark, Lyngby, Denmark.
- [14] A. Hyvarinen, J. Karhunen and E. Oja. *Independent Component Analysis*. Wiley, New York, United States, 2001.