



ESTUDO DA SEGMENTAÇÃO DO SINAL DE TÍTULOS MUSICAIS PARA CLASSIFICAÇÃO AUTOMÁTICA DE GÊNEROS

Marco Aurélio Deoldoto Paulino (PIC /UEM), Yandre Maldonado e Gomes da Costa (Orientador), e-mail: marco23_aurelio@hotmail.com.

Universidade Estadual de Maringá / Centro de Tecnologia/Maringá, PR.

Ciências Exatas e da Terra

Palavras-chave: segmentação, zoneamento, *Latin Music Database*.

Resumo

A classificação automática de gêneros musicais com o uso de técnicas de reconhecimento de padrões de sinais musicais foi apresentada por George Tzanetakis e Perry Cook em 2002 e desde então há uma série de estudos nesta área. Algumas literaturas sugerem o uso de segmentação do áudio original para a recuperação de informações relacionadas à música, como o gênero musical. O presente trabalho propõe investigar a influência do uso de diferentes trechos das músicas nas taxas de acerto obtidas na classificação automática de gêneros musicais. Neste estudo, serão investigadas algumas variações tanto em relação à quantidade de segmentos extraídos quanto ao tamanho dos mesmos. Outro aspecto investigado neste artigo está relacionado ao zoneamento da imagem dos espectrogramas, que permite criar naturalmente múltiplos classificadores. Foram escolhidas novecentas músicas da base de áudio *The Latin Music Database* (LMD) divididas igualmente em dez gêneros musicais latinos. O resultado com maior taxa de acerto se deu com o uso de um segmento de quarentas segundos a partir do vigésimo primeiro segundo até o sexagésimo segundo, com a taxa de 80,22%. A partir desse mesmo trecho foi alcançado a marca de 83,89% com o uso do zoneamento e criação de múltiplos classificadores.

Introdução

Com a grande expansão da Internet, tem se presenciado uma época de grande densidade de dados, segundo (ZETTABYTE, 2015), a previsão para o final de 2016 é que o tráfego global ultrapasse a marca de um zettabyte. Tais dados por serem disponibilizados em grande escala e oriundos de todo o mundo não apresentam um padrão na forma como são organizados, tornando o seu uso complicado. Diante dessa situação, a pesquisa em



recuperação de informações multimídia vem crescendo nos últimos tempos. Por um período, a classificação e organização de dados de músicas eram feitos baseados em metas-informações textuais, como os rótulos ID3, que possui a desvantagem de depender da ação humana. Em 2002, Tzanetakis e Cook, apresentam a classificação automática de gêneros musicais como uma tarefa de reconhecimento de padrões partir de características extraídas do sinal do áudio de músicas. Dando continuidade a pesquisa nesta área foram apresentados outros trabalho na área como Li et al. (2003), Costa, C. et al.(2004) e Koerich e Poitevin(2005).

O presente trabalho visa contribuir na investigação acerca do desempenho obtido em uso de diferentes segmentos extraídos da música. Neste sentido, investigou-se o impacto de diferentes tamanhos e quantidades de segmentos das amostras de sinal extraídas das músicas para o desenvolvimento do processo de classificação. Adicionalmente, diferentes padrões de zoneamento da imagem de espectrogramas e o impacto dos mesmos na taxa de acerto alcançada foram investigados.

Materiais e métodos

A base de dados utilizada foi a *Latin Musica Database*, no total foram utilizadas novecentas músicas divididas igualmente em dez gêneros (Axé, Bachata, Bolero, Forró, Gaúcha, Merengue, Pagode, Salsa, Sertanejo e Tango). O *artist filter*, filtro que exigem no momento da separação de *folders*, que títulos musicais de mesmos artistas sejam colocados no mesmo *fold*, foi empregado para que os resultados apresentados sejam mais concisos e confiáveis. Nas estratégias de segmentação experimentadas foram tomados cuidados para evitar que efeitos como *fade in* e *fade out*, e outros ruídos sonoros presentes no sinal, por exemplo, ruídos de plateia em músicas ao vivo, pudessem prejudicar na prejudicar o desempenho do classificador.

Os procedimentos do trabalho seguiram da seguinte forma: inicialmente era feita a escolha de quais trechos das músicas e eram extraídos via o software *wavsplit*. Após a extração, com a utilização do aplicativo *Sound eXchange* (SOX) eram geradas as imagens dos espectrogramas. Em seguida, foi realizado o zoneamento do espectrograma e extração de características de textura com o método *Local Binary Pattern* (LBP). “Neste método, a textura é descrita levando-se consideração para cada pixel C, P vizinhos equidistantes considerando-se uma distância R.” (OJALA, et al., 1996).

A classificação foi realizada com máquinas de vetores de suporte mais conhecidas como SVM, acrônimo do inglês Support Vector Machine. Para a avaliação dos resultados obtidos na classificação foi adotada a matriz de confusão, que nada mais é uma tabela a partir da qual se pode observar com que intensidade uma classe é confundida com cada uma das outras classes envolvidas no problema.



Resultados e Discussão

Inicialmente, houve uma série de testes para descobrir o(s) segmento(s) com as melhores taxas de acerto, e os principais resultados estão presentes na Tabela 1, apresentada a seguir:

Descrição do(s) segmento(s)	Taxa (%)
40 segundos centrais	80,22
50 segundos centrais	78,44
60 segundos (a partir do 31 ^{os} até o 90 ^{os})	78,22
30 segundos centrais	77,56
5 segundos (11 ^{os} ao 15 ^{os}) + 20 segundos centrais + 5 segundos finais (excluindo os 10 segundos para o término)	77,33
10 segundos (21 ^{os} ao 30 ^{os}) + 10 segundos centrais + 10 segundos finais (excluindo os 20 segundos para o término)	77
Música Inteira	75,44
Música Inteira (ignorando os 20 segundos iniciais e finais)	72,78

Tabela 1 – Resultados de diferentes segmentos. Não houve zoneamento nestes testes.

A partir do melhor resultado encontrado, quarenta segundos central, foram realizados os testes para encontrar o zoneamento com melhor taxa de acertos, como mostra a Tabela 2.

Zoneamento dos espectrogramas	Taxa (%)
Dividido em 5 zonas verticalmente	82,89
Dividido em 5 zonas verticalmente e 3 zonas horizontalmente	82,44
Dividido em 3 zonas verticalmente	82,33
Dividido em 10 zonas verticalmente	82,33
Dividido em 3 zonas verticalmente e 2 zonas horizontalmente	81,89
Sem divisão	80,22
Dividido em 3 zonas horizontalmente	78,56

Tabela 2 - Resultados de diferentes zoneamentos.

A partir dos resultados, foi observado que a escolha de trechos intermediários da música acarreta em melhores resultados em relação a segmentos no início e final da música e também em relação à música completa. Quanto ao zoneamento, as divisões em zonas correspondentes a diferentes faixas de frequência resultam na melhora na taxa enquanto em zonas correspondentes a diferentes instantes da música parecem não contribuir significativamente com a melhoria das taxas de acertos.



Conclusões

A realização deste trabalho permitiu constatar que o uso de segmentos extraídos de trechos apropriados do sinal da música pode impactar significativamente nos resultados obtidos em tarefas de classificação. Para o futuro, a pesquisa de maneiras de escolha da segmentação dinamicamente pode tornar as taxas de acertos mais altas, pois no melhor resultado, 82,89%, o gênero Forró obteve uma taxa de acertos de 72,22%, e obteve a taxa de 80% na utilização do segmento de sessenta segundos iniciais com a exclusão de trinta segundos do começo da música. Com a escolha dinâmica, a taxa de acertos por gênero aumentaria assim, aumentando a taxa geral.

Agradecimentos

Agradeço ao meu orientador Yandre, ao grupo de pesquisa e ao Programa de Educação Tutorial(MEC/SESu) pelo fomento e pelo apoio institucional.

Referências

COSTA, C.; VALLE Jr, J.; KOERICH, A. **Automatic classification of áudio data**. In: *IEEE International Conference on Systems, Man and Cybernetics*. 2004.

KOERICH, A.; POITEVIN, C. **Combination of Homogeneous Classifiers for Musical Genre Classification**. In: *IEEE International Conference on Systems, Man and Cybernetics*. 2005.

LI, T.; OGIHARA, M.; LI, Q. **A comparative study on content-based music genre classification**. In: *Proceedings of the 26th annual international ACM SIGIR Conference on research and development in information retrieval*, p. 282-289. 2003.

OJALA, T.; PIETIKÄINEN, M.; HARWOOD, D. **A comparative study of texture measures with classification based on featured distributions**. *Patter Recognition*. 1996.

THE Zettabyte Era: Trends and Analysis. 2015. Disponível em: <http://www.cisco.com/c/en/us/solutions/collateral/service-provider/visual-networking-index-vni/VNI_Hyperconnectivity_WP.html>. Acesso em: 08 jun. 2015.

TZANETAKIS, G.; COOK, P. **Musical genre classification of áudio signals**. In: *IEEE Transactions on speech and áudio processing*. 2002.