

# Classificação de Gêneros Musicais Utilizando Vetores de Característica Híbridos

Carlos N. Silla Jr.<sup>1</sup>, Celso A. A. Kaestner<sup>2</sup>, Alessandro L. Koerich<sup>3,4</sup>

<sup>1</sup>School of Computing  
University of Kent  
Canterbury, Kent, UK

<sup>2</sup>Programa de Pós-Graduação em Computação Aplicada  
Universidade Tecnológica Federal do Paraná  
Curitiba, PR, Brasil

<sup>3</sup>Programa de Pós-Graduação em Informática  
Pontifícia Universidade Católica do Paraná  
Curitiba, PR, Brasil

<sup>4</sup>Departamento de Engenharia Elétrica  
Universidade Federal do Paraná  
Curitiba, PR, Brasil

cns2@kent.ac.uk, celsokaestner@utfpr.edu.br, alekoe@computer.org

**Abstract.** *Current research on automatic music genre classification has been focusing on new classification approaches based on combining information from other sources than the audio signal. The main reason is that the use of content-based approaches, i.e. using features extracted from the audio signal, seems to have reached a glass ceiling. In this work we show that it is possible to substantially improve the accuracy in classifying music genres by merging different types of content-based features. This is an interesting result as different types of content-based features aim, at a conceptual level, to capture the same type of information. We have also used a feature selection approach based on genetic algorithms to identify which types of content-based features are responsible for the predictive accuracy gain. The analysis of the results in two benchmark databases have shown that the genetic algorithm succeeds in selecting a representative subset of features without a significant loss in accuracy. It also shows that all the different types of content-based features employed are important for the significant improvement of the classification accuracy.*

**Resumo.** *As pesquisas recentes em classificação automática de gêneros musicais têm evoluído para novas abordagens de classificação baseadas na combinação da informação de outras fontes diversas do próprio sinal de áudio. A principal razão para isso é que o uso de abordagens baseadas no conteúdo, ou seja, abordagens que utilizam características extraídas diretamente do sinal de áudio, têm atingido um patamar onde os avanços realizados não vem sendo significativos. Neste trabalho mostramos que utilizando uma combinação de diferentes conjuntos de características baseadas no conteúdo é possível melhorar substancialmente a precisão de classificação de gêneros musicais. Este resultado é interessante, pois diferentes tipos de características baseadas no conteúdo visam, em um nível conceitual, capturar o mesmo tipo de informação. Para identificarmos quais tipos de características baseadas no conteúdo são*

*responsáveis pelo ganho na precisão, utilizamos um procedimento de seleção de características baseado em algoritmos genéticos. A análise dos resultados sobre duas bases de dados de referência mostra que o algoritmo genético seleciona com sucesso um subconjunto de características representativo sem uma perda significativa na precisão de classificação. Observamos também a importância de todos os tipos de características baseadas no conteúdo para um aumento significativo na precisão de classificação.*

## 1. Introdução

A classificação automática de gêneros musicais é um tema de pesquisa bastante atual e popular na comunidade de *Music Information Retrieval* (MIR) [1]. Isso se deve ao fato de que o gênero musical é um dos argumentos utilizados com mais frequência em buscas [20, 21, 22]. Ao longo da última década este problema vem recebendo uma atenção crescente, levando ao desenvolvimento de diferentes tipos de características para representar sinais musicais [11, 12, 13], bem como a diferentes abordagens de classificação para categorizar estes sinais [1, 5, 8, 9, 17, 19]. As pesquisas desenvolvidas atualmente que fazem a análise do conteúdo do sinal de áudio vêm atingindo um patamar a partir do qual grandes avanços são cada vez mais raros [1]. Para contornar esta dificuldade, diversos autores vêm propondo o uso de diferentes tipos de informações relacionadas às músicas em combinação com as características obtidas a partir do próprio sinal de áudio [1, 8, 9, 10, 19]. Algumas fontes de informação adicional que têm sido utilizadas para expandir as abordagens baseadas em conteúdo são: informação cultural obtida a partir da utilização de técnicas de mineração WEB [10]; meta-dados booleanos representando o contexto musical [9]; letras das músicas [19]; representação simbólica a partir de um sistema de transcrição [1]; e a combinação de informação cultural e características de representação simbólica [8]. Em todas estas abordagens o uso de fontes de informação adicionais tem melhorado a precisão de classificação quando comparada a utilização somente das características baseadas no conteúdo do sinal de áudio. Entretanto, as abordagens que utilizam informação cultural ou letras são normalmente dependentes de ferramentas e técnicas de processamento de linguagem natural, que podem não estar disponíveis em linguagens outras que não o inglês.

Diferentemente do que acontece em reconhecimento de voz, onde existem características estabelecidas a longo tempo, como é o caso dos coeficientes cepstrais de frequência mel ou simplesmente MFCCs, os pesquisadores que trabalham com a classificação de gêneros musicais ainda não chegaram a um consenso a respeito de quais são as características que melhor representam os diferentes gêneros musicais. Uma alternativa é a utilização de diferentes tipos de características baseadas no próprio conteúdo do sinal de áudio, concatenadas em um único vetor de características. Uma das propostas pioneiras que emprega diferentes característica em um único vetor é o ambiente MARSYAS<sup>1</sup> [11]. Em um estudo recente Lidy et al. [5] mostrou que é possível melhorar a precisão na classificação de gêneros musicais e na identificação de artistas, utilizando diferentes tipos de características baseadas no conteúdo. Os autores concatenam diferentes tipos de características em um único vetor. Entretanto, esta combinação é feita “manualmente” seguindo a intuição dos autores de que trocando algumas características específicas, em particular, as características relacionadas à batida do MARSYAS [11] por outras características rítmicas, o vetor híbrido resultante seria capaz de capturar um número maior de aspectos rítmicos e timbrais da música. Considerando que as características foram escolhidas de maneira totalmente empírica, podemos levantar algumas

---

<sup>1</sup><http://marsyas.info>

questões relevantes: seria esta a melhor maneira de se criar um vetor de características híbrido? Por que não utilizar simplesmente toda a informação disponível, isto é, todas as características baseadas em conteúdo juntas e empregar um algoritmo de seleção de características para buscar automaticamente vetor de características híbrido a partir de todas as características disponíveis?

Neste trabalho, partindo de quatro conjuntos de características considerados atualmente o estado da arte para classificação de gêneros musicais e empregados em [5], obtivemos um único vetor de características híbrido. Observamos que esta combinação dos quatro conjuntos de características leva a uma precisão de classificação substancialmente melhor do que utilizar somente qualquer um dos tipos de características individualmente. Para compreender melhor a razão pela qual o vetor híbrido completo fornece resultados tão bons em relação a vetores que utilizam individualmente os conjuntos de características considerados, empregamos um algoritmo de seleção de características baseado em algoritmos genéticos, para, de alguma forma, ganharmos algum conhecimento sobre quais tipos de características seriam as mais relevantes. Além disso, verificamos que o procedimento de seleção de características é capaz de criar automaticamente um vetor de características híbrido de dimensão reduzida que nos possibilita alcançar uma precisão similar àquela fornecida pelo vetor de características híbrido completo.

Apesar de alguns trabalhos recentes terem empregado técnicas de seleção de características na classificação de gêneros musicais, eles focaram na utilização de diferentes técnicas visando a redução na dimensionalidade do espaço de características [6] ou buscando uma possível melhora na precisão de classificação [7]. Neste trabalho escolhemos um procedimento baseado em algoritmos genéticos que além de buscar a redução na dimensionalidade e melhora na precisão, torne possível também analisar a importância dos diferentes tipos de conjuntos de características utilizadas pelo vetor híbrido completo.

Este artigo está organizado como se segue. A Seção 2 apresenta uma visão geral da abordagem proposta. A Seção 3 apresenta de maneira sucinta uma descrição dos conjuntos de características utilizados neste trabalho. A Seção 4 apresenta os detalhes do algoritmo de seleção de características utilizado neste trabalho. Os resultados experimentais obtidos sobre duas bases de dados distintas são apresentados na Seção 5. As conclusões e perspectivas de trabalhos futuros são apresentados na última seção.

## 2. Visão Geral do Sistema

Neste trabalho utilizamos a abordagem de decomposição temporal [16, 17] com um vetor de características híbrido completo com duzentos e noventa e oito dimensões que agrega as características dos quatro conjuntos, sendo quarenta dimensões relativas aos coeficientes do histograma do intervalo Inset-Onset [12], trinta dimensões do MARSYAS [11], sessenta dimensões dos histogramas de ritmo [13] e cento e sessenta e oito dimensões dos descritores estatísticos do espectro [13]. Para cada música todas estas características são extraídas de três segmentos de trinta segundos das partes inicial (I), do meio (M) e do final (F) do sinal de áudio, como ilustrado na Figura 1.

Dado o vetor de características híbrido, um procedimento de seleção de características baseado em algoritmos genéticos é aplicado sobre este vetor. Designamos como *AG-Sel*, o vetor de características resultante do procedimento de seleção de características. O procedimento de seleção de características é aplicado individualmente aos vetores de características extraídos de cada segmento da música (I, M, F). Para cada segmento, um classificador multi-classes baseado em máquinas de vetor de suporte (SVM) é treinado utilizando uma metodologia 1x1. Os classificadores SVM utilizam o método de otimização sequencial mínimo e um kernel Gaussiano.

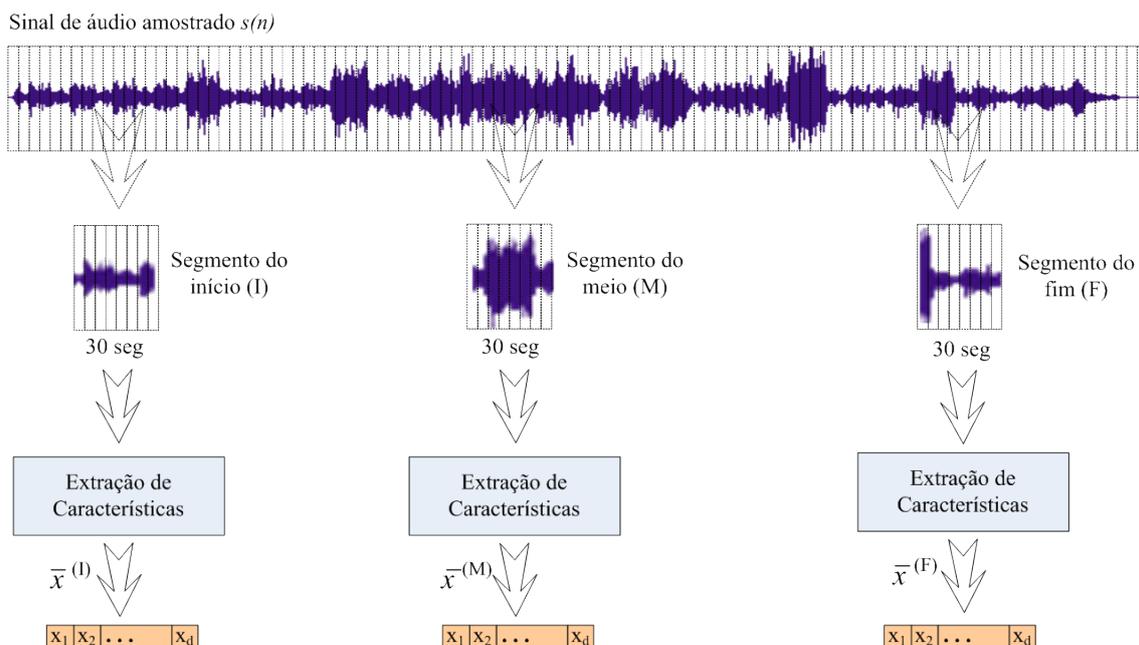


Figura 1: Extração de características

### 3. Conjunto de Características

Nesta seção descrevemos de maneira sucinta os quatro diferentes conjuntos de características que são utilizados para criar o vetor de características híbrido: coeficientes do histograma de intervalos Inset-Onset, MARSYAS, histogramas de ritmo e descritores estatísticos de espectro.

#### 3.1. Coeficientes do Histograma de Intervalos *Inset-Onset* (IOIHC)

Neste conjunto, as características estão relacionadas às propriedades rítmicas do sinal musical, mais especificamente ao tempo entre o início dos pontos de ataque de eventos sucessivos ou notas [12, 4]. O conjunto de características é computado a partir de uma função particular de periodicidade rítmica que representa a saliência normalizada com respeito ao período dos intervalos *inset-onset* presentes no sinal. Os valores de características são então parametrizados da seguinte maneira: (a) projeção do eixo de período inset-onset da escala linear para a escala Mel, que possui uma dimensionalidade menor, através de uma filtragem; (b) cálculo do logaritmo da magnitude do histograma inset-onset; (c) cálculo da transformada inversa de Fourier, mantendo os primeiros quarenta coeficientes. Estes passos produzem um conjunto de características que é análogo aos MFCCs, mas no domínio de períodos rítmicos em vez de frequências do sinal. Uma descrição mais detalhada destas características pode ser encontrada em [12]. O conjunto de características final é representado por um vetor de características de quarenta dimensões.

#### 3.2. MARSYAS

Este conjunto de características foi proposto por Tzanetakis e Cook [11]. O conjunto de características original pode ser dividido em três subconjuntos: características do conteúdo rítmico, características descrevendo a textura timbral e características relacionadas ao tom. Características relacionadas ao ritmo visam representar a regularidade do ritmo e as saliências e períodos relativos dos diversos níveis da hierarquia métrica. Elas são baseadas no histograma de batidas que é uma função particular da periodicidade rítmica que representa a força da batida. As características são calculadas a partir de algumas medidas estatísticas obtidas a partir do histograma (amplitudes relativas, períodos,

razões entre picos salientes, soma do histograma como uma indicação da força da batida). As características de textura timbral são obtidas a partir da transformada de Fourier de tempo reduzido (STFT) do sinal de áudio e inclui a média e a variância do centróide espectral com janelamento, decaimento do sinal com a frequência (*rolloff*), fluxo espectral, taxa de cruzamento por zero, baixa energia e os cinco primeiros coeficientes ceptrais de frequência Mel. Finalmente, as características relacionadas ao tom incluem os períodos máximos do pico do tom a partir de histogramas de tom. O conjunto de características completo pode ser diretamente obtido do ambiente MARSYAS<sup>2</sup>, que implementa todos os procedimentos necessários. O vetor de características final é de trinta dimensões.

### 3.3. Histogramas de Ritmo (RH)

Este conjunto de características é baseado em modelos psicoacústicos que capturam ritmos e outras flutuações nas faixas de frequência críticas ao sistema auditivo humano [13, 1]. O conjunto de características é obtido da seguinte maneira: (1) a sensação específica de volume e intensidade nas vinte e quatro bandas de frequência críticas são calculadas através da transformada de Fourier de tempo reduzido (STFT); (2) as bandas de frequência resultantes são agrupadas na escala Bark, aplicando funções de espalhamento para levar em conta efeitos de mascaramento. São aplicadas sucessivas transformações nas escalas Decibel, Phon e Sone. A escala Bark é uma escala perceptual que agrupa frequências para bandas críticas de acordo com regiões perceptivas de tom [23]; estes passos produzem uma representação sonográfica psicoacústica modificada que reflete a sensação humana de volume e intensidade; (3) a transformada discreta de Fourier é aplicada ao sonograma resultando em um espectro variante no tempo da modulação de amplitude do volume e intensidade por frequência de modulação para cada banda crítica individual; (4) um procedimento de redução dimensional é aplicado, onde os valores de amplitude de modulação são agrupados em vinte e quatro bandas críticas individuais, exibindo a magnitude para sessenta frequências de modulação desde 0,17 a 10 Hz.

O procedimento completo é aplicado sucessivamente a diversas amostras do sinal de áudio, de modo que o vetor de características final é calculado como a mediana dos valores para cada amostra de áudio. O vetor de características final possui sessenta dimensões.

### 3.4. Descritores Estatísticos de Espectro (SSD)

Este conjunto de características é obtido a partir de algumas sensações específicas de volume e intensidade da música [1]. A sensação de volume e intensidade é computada em vinte e quatro bandas da escala Bark. Medidas estatísticas que incluem média, mediana, variância, assimetria, curtose, e valores de mínimo e máximo são computados para cada banda. O conjunto de características obtido representa as flutuações nas bandas críticas e também captura informação timbral adicional que não é abrangida pelos conjuntos de características prévios. O vetor de características final possui cento e sessenta e oito dimensões.

## 4. Seleção de Características

Em diversos problemas de reconhecimento de padrões é difícil encontrar um conjunto adequado de características. Procedimentos de seleção de características tentam escolher um subconjunto a partir de um conjunto de características original visando reduzir o esforço computacional nas etapas de pré-processamento e classificação tentando, porém conservar ou mesmo aumentar a precisão de classificação [2, 15]. Em geral, este problema

---

<sup>2</sup><http://marsyas.info>

pode ser modelado como um problema de busca heurística, onde os pontos no espaço de busca são todos os subconjuntos possíveis que podem ser gerados a partir de um conjunto de características original, e a heurística que avalia as soluções é a precisão fornecida pelo algoritmo de classificação.

Os procedimentos para seleção de características podem se basear em duas ideias principais [3]: (a) na abordagem filtro o processo de seleção de características é executado como um passo pré-processamento independente, antes da aplicação de qualquer algoritmo de classificação; (b) na abordagem *wrapper* o algoritmo de classificação é empregado como uma subrotina pelo procedimento de seleção de características, trabalhando especificamente para avaliar as soluções geradas. É importante notar que utilizamos vetores de características gerados a partir de segmentos extraídos do início, do meio e do final do sinal de áudio. Logo, o procedimento de seleção de características é aplicado independentemente em cada um destes segmentos do sinal de áudio. Portanto, é possível também analisar a relativa importância das características de acordo com a sua origem no sinal de áudio.

Utilizamos um paradigma clássico de algoritmos genéticos para realizar a busca, pois este algoritmo é reconhecidamente um método de busca eficiente em problemas complexos, como é o caso de buscar um subconjunto de características ótimo a partir de um vetor de alta dimensão. Em nosso arcabouço, os indivíduos (ou cromossomos) são associados a vetores com atributos binários de  $d$  dimensões, onde  $d$  é o tamanho máximo do vetor. Cada indivíduo atua como uma máscara binária para um vetor de características associado, isto é, o valor 1 indica que a dimensão correspondente deve ser utilizada, enquanto o valor 0 indica que ela deve ser descartada. A função de avaliação dos indivíduos é calculada diretamente a partir da precisão de classificação do respectivo classificador, de acordo com a abordagem *wrapper*.

Resumimos os passos do algoritmo genético de seleção de características utilizado como se segue:

1. Gerar aleatoriamente uma população inicial;
2. Até que o processo de evolução convirja (não haja uma mudança significativa na população entre duas gerações consecutivas OU o número máximo de iterações seja atingido) FAÇA:
  - (a) PARA cada indivíduo na população FAÇA:
    - i. Treine o classificador usando as características selecionadas (de acordo com a máscara de bits);
    - ii. Aplique a estrutura de classificação obtida a um conjunto de validação para calcular a precisão, que é atribuída à função de avaliação do indivíduo;
  - (b) Aplique o elitismo para conservar os indivíduos bem posicionados E aplique operadores de cruzamento e mutação para obter a próxima geração;
3. Os indivíduos bem posicionados determinam o conjunto de características final selecionado.

## 5. Resultados Experimentais

Os experimentos descritos nesta seção foram realizados utilizando duas bases de dados de referência para a classificação de gêneros musicais. A primeira base de dados é a Latin Music Database (LMD) [18] que contém 3.227 músicas de dez gêneros musicais Latinos, sendo 313 do gênero Axé, 313 do gênero Bachata, 315 do gênero Bolero, 313 do gênero Forró, 311 do gênero Gaúcha, 315 do gênero Merengue, 307 do gênero Pagode, 311 do

gênero Salsa, 321 do gênero Sertaneja e 408 do gênero Tango. A segunda base de dados é a ISMIR Genre 2004 [14] que contém 1.458 músicas de seis gêneros musicais populares, sendo 640 do gênero clássica, 229 do gênero eletrônica, 52 do gênero jazz&blues, 90 do gênero metal&punk, 203 do gênero rock&pop e 244 do gênero world. Ambas bases de dados vêm sendo utilizadas na competição internacional *The Music Information Retrieval Evaluation eXchange (MIREX)*<sup>3</sup>.

Todos os experimentos descritos nesta seção utilizaram um procedimento de validação cruzada estratificada de ordem dez. As Tabelas 1 e 2 apresentam os resultados sobre a Latin Music Database (LMD). A primeira coluna das tabelas relacionam os diferentes conjuntos de características avaliados neste trabalho. A terceira linha das tabelas apresenta os resultados para o conjunto de características descritores estatísticos de espectro (SSD). A quarta linha apresenta a precisão para o conjunto de características SSD-Híbrido, isto é, um conjunto de características criado “manualmente” através da concatenação das características SSD e MARSYAS, excluindo, porém, as características relacionadas à batida do MARSYAS [5]. A quinta linha apresenta os resultados utilizando todas as características concatenadas em um único vetor (ALL), ou seja, a concatenação das características IOIHC, MARSYAS, RH e SSD. A última linha da tabela apresenta os resultados obtidos a partir do vetor de características híbrido (ALL) após a seleção de características feita pelo algoritmo genético (ALL-Sel). Na Tabela 1 os resultados são apresentados para cada segmento da música, ou seja, para vetores extraídos do início (I), meio (M) e final (F) da música, enquanto que na Tabela 2 são apresentados os resultados obtidos pela combinação dos classificadores aplicados aos segmentos de início (I), meio (M) e final (F) da música, usando as regras votação da maioria (MAJ), soma (SUM) e produto (PROD) [17].

A significância estatística dos resultados foi medida pelo teste t-Student pareado bicaudal usando um nível de confiança de 95% para verificar se os resultados obtidos usando um vetor com todas as características (ALL) foram significativamente melhores do que os resultados obtidos a partir dos vetores considerando subconjuntos. Nos casos onde a diferença é estatisticamente significante, o símbolo “\*” é incluído. Além disso, apresentamos somente os resultados para SSD e SSD-Híbrido, pois estes conjuntos de características individuais foram os que propiciaram os melhores desempenhos em um estudo anterior [5].

Características	Classificação Correta (%)		
	Segmento		
	I	M	F
SSD	74.70* ± 2.25	81.86* ± 1.74	79.93* ± 2.08
SSD-Híbrido [5]	78.76* ± 2.26	85.30* ± 1.39	82.80* ± 1.13
ALL	81.66 ± 1.52	88.03 ± 2.00	84.06 ± 1.09
ALL-Sel	80.40* ± 2.51	87.40 ± 1.76	84.13 ± 1.57

**Tabela 1: Precisão (%) e desvio padrão (±) sobre a Latin Music Database (LMD) para classificadores individuais**

A análise dos resultados das Tabelas 1 e 2 mostra que para a LMD o vetor de características completo (ALL) propicia resultados significativamente melhores do que os resultados obtidos com a utilização somente das características SSD ou SSD-Híbrido em todos os casos (segmentos individuais e combinados com regras de classificação). Este é um resultado interessante, pois estes conjuntos de características frequentemente

<sup>3</sup><http://www.music-ir.org/mirex/>

Características	Classificação Correta (%)		
	Regra de Combinação		
	MAJ	SUM	PROD
SSD	83.93* ± 1.52	83.30* ± 1.32	82.56* ± 1.75
SSD-Híbrido [5]	87.53* ± 1.20	87.40* ± 1.01	87.06* ± 1.15
ALL	89.53 ± 1.26	88.93 ± 1.19	88.76 ± 1.11
ALL-Sel	88.80* ± 1.17	88.33 ± 1.22	88.16 ± 1.25

**Tabela 2: Precisão (%) e desvio padrão (±) sobre a Latin Music Database (LMD) para a combinação dos classificadores**

obtiveram os melhores resultados quando comparados aos outros conjuntos de características utilizados no vetor de características completo (IOIHC , MARSYAS, RH) [5]. Quando comparado ao vetor híbrido com as características selecionadas (ALL-Sel), o vetor de características completo (ALL) fornece resultados significativamente melhores em apenas dois dos seis casos.

As Tabelas 3 e 4 apresentam os resultados para a base de dados ISMIR Genre 2004. Cada linha desta tabela tem significado similar às linhas das Tabelas 1 e 2, descritas anteriormente.

Características	Classificação Correta (%)		
	Segmento		
	I	M	F
SSD	71.20* ± 2.52	76.12* ± 3.76	72.71* ± 2.42
SSD-Híbrido [5]	75.40* ± 3.03	79.00 ± 3.13	76.07 ± 2.94
ALL	77.08 ± 3.41	78.74 ± 2.59	76.91 ± 2.44
ALL-Sel	76.11 ± 4.11	80.27* ± 2.53	77.51 ± 2.39

**Tabela 3: Precisão (%) e desvio padrão (±) sobre a base de dados ISMIR GENRE 2004 para classificadores individuais**

Características	Classificação Correta (%)		
	Regra de Combinação		
	MAJ	SUM	PROD
SSD	77.20* ± 3.28	76.52* ± 2.57	76.73* ± 2.62
SSD-Híbrido [5]	80.37 ± 2.65	79.75 ± 2.65	79.71* ± 2.71
ALL	81.51 ± 2.76	82.14 ± 2.50	81.87 ± 2.53
ALL-Sel	81.51 ± 2.12	82.37 ± 2.46	82.43 ± 2.77

**Tabela 4: Precisão (%) e desvio padrão (±) sobre a base de dados ISMIR GENRE 2004 para a combinação dos classificadores**

Uma análise dos resultados das Tabelas 3 e 4 indica que o uso do vetor de características híbrido (ALL) propicia uma precisão de classificação significativamente melhor do que as obtidas quando utilizamos o vetor somente com o conjunto de características SSD (que nos experimentos prévios tiveram desempenho melhor do que qualquer outro conjunto de características nesta base de dados) em todos os casos. Quando comparado com o conjunto de características SSD-Híbrido, o vetor de características híbrido (ALL) fornece melhores resultados em cinco dos seis casos. Contudo, somente em dois casos a diferença é estatisticamente significativa. O vetor híbrido com as características selecionadas pelo algoritmo genético (ALL-Sel) propicia uma precisão maior do que o vetor

de características completo (ALL) em cinco (sendo um estatisticamente significativo) dos seis casos.

Como estamos usando um algoritmo de seleção de características, além da precisão de classificação, existem outros aspectos que merecem uma análise. Primeiro, se o algoritmo genético foi capaz de selecionar um subconjunto reduzido de características. Segundo, queremos saber quais características foram selecionadas dentre todas disponíveis em todos os conjuntos. Existe algum tipo de característica que não foi selecionada? Existe uma relação entre as características utilizadas no método manual que é composto por um conjunto de características em particular e as características do MARSYAS (excluindo as características relacionadas a batida) e as características selecionadas automaticamente pelo algoritmo genético?

A Tabela 5 mostra o número de características para cada representação, para cada segmento e para as bases de dados LMD e ISMIR. A análise da Tabela 5 revela que o procedimento baseado em algoritmos genéticos para criar de maneira automática vetores de características híbridos seleciona até 62,40% das características presentes no vetor de características completo (ALL). O número de características selecionadas é também menor que o vetor de características híbrido manualmente selecionado (SSD-Híbrido) e, como visto, também fornece melhor precisão de classificação.

Características	Número de Características		
	Segmento		
	I	M	F
SSD	168	168	168
SSD-Híbrido [5]	192	192	192
ALL	298	298	298
ALL-Sel (LMD)	171	186	179
ALL-Sel (ISMIR)	163	168	163

**Tabela 5: Número de características dos vetores para cada conjunto/subconjunto avaliado em cada segmento**

A Tabela 6 apresenta o número de características selecionadas de cada conjunto de características em particular para compor o vetor de características híbrido gerado pelo algoritmo genético. A análise desta tabela mostra que as características foram selecionadas de todos os quatro tipos de conjuntos. Este resultado implica que o vetor híbrido de características baseadas no conteúdo se beneficia de todos os diferentes conjuntos.

Características	Base de Dados					
	ISMIR			LMD		
	Segmento			Segmento		
	I	M	F	I	M	F
IOIHC	18	18	23	25	23	20
MARSYAS	20	17	22	16	25	18
RH	34	33	24	36	32	38
SSD	91	100	94	94	106	103

**Tabela 6: Número de características selecionadas pelo algoritmo genético de cada conjunto de características, para cada segmento, para compor o vetor ALL-Sel**

## 6. Conclusões

Na última década muitos autores propuseram diferentes descritores para músicas ou maneiras de combinar estes descritores com informações de outras fontes (letras, informação cultural, características simbólicas, etc). Entretanto, um aspecto que vem sendo negligenciado e que foi investigado neste trabalho, é que a junção de diferentes tipos de características baseadas em conteúdo, é possível alcançar uma melhor precisão de classificação em relação a utilização de características individualmente. Os resultados obtidos são também melhores que vetores formados a partir de características “manualmente” selecionadas.

Utilizamos também seleção de características baseada em um algoritmo genético para criar um vetor de características híbrido de dimensão reduzida que fornece precisão similar ao vetor de características contendo todos os tipos de características, mas com uma representação mais compacta. Verificamos também que todos os tipos de representação de características têm um papel importante na melhora da precisão de classificação. Analisamos também quais características de cada conjunto de características e verificamos que todos os tipos de diferentes características possuem um papel importante na melhoria da precisão de classificação, tendo em vista que o algoritmo genético sempre selecionou características de todos os quatro tipos de conjuntos utilizados neste trabalho.

Quando são comparados o conjunto de características selecionado pelo algoritmo genético e o conjunto de características selecionado por humanos, o primeiro seleciona características relacionadas à batida do MARSYAS. Este resultado é igualmente interessante, pois a justificativa teórica para o vetor de características híbrido selecionado por humanos em [5] é de que outros descritores de características forneceriam informação mais precisa do que as características baseadas na batida do MARSYAS. Parece que na prática os diferentes tipos de características podem complementar de maneira eficiente umas às outras.

Como trabalho futuro pretendemos realizar experimentos controlados para verificar o impacto real de fontes de informação adicional em relação ao vetor híbrido de características baseadas no conteúdo utilizado neste trabalho. Isto é particularmente importante, pois a maioria das abordagens que utilizam fontes adicionais de informação depende de ferramentas de processamento de linguagem natural, o que para línguas diversas do Inglês, podem não estar disponíveis.

## Referências

- [1] T. Lidy, A. Rauber, A. Pertusa and J. M. Iñesta. Improving Genre Classification by Combination of Audio and Symbolic Descriptors Using a Transcription System. *Proc. 8th International Conference on Music Information Retrieval*, Vienna, Austria, September 23–27, 2007
- [2] A. Blum and P. Langley. Selection of Relevant Features and Examples in Machine Learning, *Artificial Intelligence* **97** (1997) 245–271.
- [3] L. C. Molina, L. Belanche and A. Nebot. Feature Selection Algorithms: a Survey and Experimental Evaluation, in *Proc. of the IEEE Int. Conference on Data Mining*, Maebashi City, JP, 2002, pp. 306–313.
- [4] F. Gouyon, P. Herrera and P. Cano. Pulse-dependent Analysis of Percussive Music, in *Proc. of the 22th Int. AES Conference on Virtual, Synthetic and Entertainment Audio*, Espoo, Finland, 2002.

- [5] T. Lidy, C. N. Silla Jr., O. Cornelis, F. Gouyon, A. Rauber, C. A. A. Kaestner, A. L. Koerich. On the Suitability of State-of-the-Art Music Information Retrieval Methods for Analyzing, Categorizing, Structuring and Accessing Non-Western and Ethnic Music Collections. *Signal Processing*, Vol.90 No.4, pp.1032–1048.
- [6] M. Grimaldi, P. Cunningham and A. Kokaram. A Wavelet Packet Representation of Audio Signals for Music Genre Classification using Different Ensemble and Feature Selection Techniques, in *Proc. 5th ACM SIGMM Int. Workshop on Multimedia Information Retrieval*, Berkeley, CA, USA, 2003, pp. 102–108.
- [7] R. Fiebrink and I. Fujinaga. Feature Selection Pitfalls and Music Classification, in *Proc. of the 7th Int. Conf. on Music Information Retrieval (ISMIR06)*, Victoria, CA, USA, 2006, pp. 340–341.
- [8] C. McKay and I. Fujinaga. Combining Features Extracted from Audio, Symbolic and Cultural Sources. *Proc. 9th International Conference on Music Information Retrieval*, Philadelphia, USA, 2008, pp. 597–602.
- [9] J. J. Aucouturier, F. Pachet, P. Roy and A. Beurivé, Signal + Context = Better Classification. *Proc. 8th International Conference on Music Information Retrieval*, Vienna, Austria, 2007, pp. 425–430.
- [10] B. Whitman and P. Smaragdis, Combining Musical and Cultural Features for Intelligent Style Detection. *Proc. Intl. Conf. on Music Information Retrieval*, 2002, pp. 47–52.
- [11] G. Tzanetakis and P. Cook. Musical Genre Classification of Audio Signals. *IEEE Transactions on Speech and Audio Processing*, 2002, 10, pp. 293-302.
- [12] F. Gouyon, S. Dixon, E. Pampalk and G. Widmer, Evaluating Rhythmic Descriptors for Musical Genre Classification. *Proc. 25th Intl. AES Conference*, 2004, London, UK.
- [13] T. Lidy and A. Rauber. Evaluation of Feature Extractors and Psycho-acoustic Transformations for Music Genre Classification. *Proc. 6th Intl. Conf. on Music Information Retrieval*, 2005, pp. 34-41, London, UK.
- [14] P. Cano, E. Gómez, F. Gouyon and P. Herrera, M. Koppenberger, B. Ong, X. Serra, S. Streich and N. Wack. ISMIR 2004 Audio Description Contest, Music Technology Group, Pompeu Fabra University, 2006, Technical Report MTG-TR-2006-02.
- [15] M. Dash and H. Liu. Feature Selection for Classification, *Intelligent Data Analysis 1* (1997) 131–156.
- [16] C. N. Silla Jr., C. A. A. Kaestner and A. L. Koerich. Automatic Music Genre Classification Using Ensemble of Classifiers, *Proc. IEEE Intl. Conf. on Systems, Man and Cybernetics*, 2007, pp. 1687–1692.
- [17] C. N. Silla Jr., A. L. Koerich and C. A. A. Kaestner. A Machine Learning Approach to Automatic Music Genre Classification, *Journal of the Brazilian Computer Society*, 2008, 14 (3), 7–18.
- [18] C. N. Silla Jr., A. L. Koerich, and C. A. A. Kaestner. The Latin Music Database. *Proc. 9th International Conference on Music Information Retrieval*, 2008, pp. 451-456.
- [19] R. Mayer, R. Neumayer and A. Rauber. Combination of Audio and Lyrics Features for Genre Classification in Digital Audio Collections, *Proc. ACM Intl. Conf. on Multimedia*, 2008, 159-168.

- [20] J. S. Downie and S. J. Cunningham. Toward a theory of music information retrieval queries: system design implications. *Proc. 3rd International Conference on Music Information Retrieval*, pp. 299-300, 2002.
- [21] J. H. Lee and J. S. Downie. Survey of music information needs, uses, and seeking behaviours preliminary findings. *Proc. 5th International Conference on Music Information Retrieval*, Barcelona, Spain, pp. 441-446, 2004.
- [22] E. Pampalk, A. Rauber and D. Merkl. Content-Based organization and visualization of music archives. *Proc. ACM Multimedia*, Juan-les-Pins, France, pp. 570-579, 2002.
- [23] E. Zicker and H. Fastl. *Psychoacoustics - Facts and Models*, Springer Series of Information Sciences **22**, Springer, Berlin, 1999.