CARLOS NASCIMENTO SILLA JUNIOR

COMBINAÇÃO DE CLASSIFICADORES PARA O RECONHECIMENTO AUTOMÁTICO DE GÊNEROS MUSICAIS

Dissertação apresentada ao Programa de Pós-Graduação em Informática da Pontifícia Universidade Católica do Paraná como requisito parcial para obtenção do título de Mestre em Informática.

Curitiba
2007

CARLOS NASCIMENTO SILLA JUNIOR

COMBINAÇÃO DE CLASSIFICADORES PARA O RECONHECIMENTO AUTOMÁTICO DE GÊNEROS MUSICAIS

Dissertação apresentada ao Programa de Pós-Graduação em Informática da Pontifícia Universidade Católica do Paraná como requisito parcial para obtenção do título de Mestre em Informática.

Área de Concentração: Ciência da Computação

Orientador: Celso A. A. Kaestner
Co-orientador: Alessandro L. Koerich

Curitiba
2007

Silla Junior, Carlos Nascimento
COMBINAÇÃO DE CLASSIFICADORES PARA O RECONHECIMENTO AUTOMÁTICO DE GÊNEROS MUSICAIS. Curitiba, 2007.

Dissertação - Pontifícia Universidade Católica do Paraná. Programa de Pós-Graduação em Informática.

1. Classificação Automática de Gêneros Musicais 2. Combinação de Classificadores 3. Seleção de Atributos I.Pontifícia Universidade Católica do Paraná. Centro de Ciências Exatas e Tecnologia. Programa de Pós-Graduação em Informática II - t

Dedico este trabalho à memória do meu amado avô Eymar, à minha amada avó Alba e aos meus amados pais Ana e Carlos.

Agradecimentos

A Deus pela vida e pela saúde.

À minha avó por quem palavras não são suficientes para agradecer todo o amor, carinho e incentivo.

À minha mãe por ser, além da melhor mãe do mundo, o meu maior exemplo de força e determinação na vida.

Ao meu pai por compartilhar toda sua sabedoria e experiência de vida.

À minha namorada por todos os momentos incríveis que passamos juntos e pela compreensão nos momentos em que tive que trabalhar na dissertação.

Ao meu amigo e orientador Celso Kaestner por todos os anos em que tive o privilégio de ser orientado por ele.

Ao meu amigo e co-orientador Alessandro Koerich pelas críticas valiosas e construtivas.

Aos meus mais que amigos, Luiz, Aron, Marcia, Kelly, e Osmar, por todos os momentos em que estiveram ao meu lado, bons ou ruins.

Aos meus amigos e professores do Centro de Dança Jaime Aroxa - Paraná.

Aos meus colegas, Sandra, Yuri, Luciane, Fernanda e Leandro, e ao meu orientador Antônio dos Santos Neto do curso de especialização em comunicação e semiótica que sempre me incentivaram a não abandonar a pós em função do mestrado.

Aos meus amigos, colegas e professores do PPGIA que tornaram os últimos anos não só instrutivos como também divertidos.

À Pontifícia Universidade Católica do Paraná pela oportunidade de continuar os meus estudos em nível de especialização e mestrado.

Sumário

Lista de Tabelas

Lista de Figuras

Lista de Símbolos

k	Número de vizinhos mais próximos do classificador k-NN
nc	Número de Classes
C_i	i-ésima classe
x	Vetor de características
p	Probabilidade ou escore de confiança
$\hat{C}$	Classe final
$p(C_i\|x)$	Probabilidade a *posteriori* da classe C_i dado x
l	Número de classificadores binários utilizados pelo método de Round Robin
e	Novo exemplo
M	Número de vetores de características
r	Um rótulo possível
R	Conjunto de rótulos possíveis
nf	Número de características escolhidas pelo algoritmo de seleção
D	Número de características do conjunto original
S	Sinal de áudio de uma música
g	Um gênero musical
$\mathcal{G}$	Conjunto de todos os gêneros musicais
$s(i)$	Sinal amostrado no instante i
A	Número total de amostras que formam o sinal de áudio digital
f	Freqüencia de amostras por segundo
t_w	Duração em segundos de uma janela de extração de características
t_m	Duração em segundos do intervalo para extração de características
ft_w	Número de amostras de áudio em cada segmento
q	Número de quadros existentes na música
t_{w_i}	Ponto inicial de um segmento
$M_t[n]$	Valor da Transformada de Fourrier no quadro t e faixa de freqüência n
t	Quadro
n	Faixa de Freqüência

Lista de Abreviações

MIR	*Multimidia Information Retrieval*
RP	*Reconhecimento de Padrões*
k-NN	*k Vizinhos Mais Próximos*
OAA	*Um-Contra-Todos (One-Against-All)*
RR	*Round Robin*
MAJ	*Voto da Maioria*
MAX	*Máximo*
SUM	*Soma*
WS	*Soma Ponderada*
PROD	*Produto*
WP	*Produto Ponderado*
AGs	*Algoritmos Genéticos*
MFCC	*Coeficientes Cepstrais de Freqüência-Mel*
DWCH	*Histogramas de Coeficientes fornecidos pela Daubechies Wavelet*
LDA	*Análise discriminante linear*
SVM	*Máquinas de suporte vetorial*
J48	*Árvores de Decisão*
NB	*Naive Bayes*
RSM	*Método de Busca em Subespaços Aleatórios*
PCA	*Análise de Componentes Principais*
IG	*Ganho de Informação*
GR	*Razão de Ganho*
DWPT	*Transformada Wavelet Discreta*
MLP	*Rede Neural do tipo Multi-Layer Perceptron*
LNN	*Rede neural simples de uma camada*
LDC	*Linear classifier assuming normal densities with equal covariance matrices*
QDC	*Quadratic classifier assuming normal densities*
UDC	*Quadratic classifier assuming normal uncorrelated densities*

GTZAN	*Base de dados desenvolvida no trabalho de Tzanetakis e Cook (2002)*
FFS	*Seleção de características com busca para frente*
BFS	*Seleção de características com busca para trás*
ACE	*Autonomous Classifier Engine*
OMEN	*On demand Metadata ExtractioN*
SAL	*Song Artist List*
MFCC	*Coeficientes Cepstrais de Freqüência-Mel*
STFT	*Short Time Fourrier Transform*
BPM	*Batidas Por Minuto*
MI	*Música Inteira*
BL	*Baseline*

Resumo

Dentro do contexto das tarefas de reconhecimento de padrões, uma tarefa recente é a classificação automática de gêneros musicais. Nessa perspectiva a música constitui um objeto interessante de estudo, pois ela pode ser representada como um sinal contínuo que varia no tempo. Neste trabalho é apresentado um novo método para a classificação automática de gêneros musicais utilizando combinação de classificadores baseada nos métodos de segmentação do sinal de áudio (*Time Decomposition*), na decomposição do espaço do problema (*Space Decomposition*) e em regras de combinação que utilizam a probabilidade a posteriori dos classificadores. Também são avaliados métodos de seleção de características visando reduzir o esforço computacional necessário durante a extração das características do sinal de áudio das melodias das peças músicas no intuito de verificar se todas as características utilizadas são realmente importantes, ou se é possível utilizar apenas um subconjunto das mesmas, dessa forma reduzindo o número de características a serem computadas a partir do sinal de áudio. Além disso, foi desenvolvida uma nova base de dados para o problema, denominada *Latin Music Database*, que contém 3.160 músicas de 10 gêneros latinos. Os resultados experimentais mostram que o método proposto permite classificar corretamente o gênero de 66.76% das músicas. Isso representa uma melhora de 9.33% em relação ao método tradicional utilizando o melhor classificador individual que considera o uso de apenas um único trecho do início da música. Utilizando um vetor com um número reduzido de características foi obtida uma melhora de 7.43% ao utilizar o método de combinação.

Palavras-chave: Classificação Automática de Gêneros Musicais, Combinação de Classificadores, Seleção de Atributos, Reconhecimento de Padrões de Sinais de Áudio.

Abstract

In the Pattern Recognition (PR) area, a subject that is recently getting attention is the automatic music genre classification. From the PR perspective, music recordings are an interesting object for study as they are a continuous signal that vary over time. In this work a novel approach for the task of automatic music genre recognition is presented. This approach is based on the ensemble of classifiers based on the time segmentation of the audio signal (*Time Decomposition*), the methods for problem *Space Decomposition* and in combination rules that uses the a posteriori probability given by the classifiers. Feature selection methods are also evaluated in an effort to reduce the computational cost of the feature extraction phase and also in order to verify wether all the features are important to the task, or if it is possible to use only a subset of them. Another important aspect of this work is that a new database containing 3.160 music pieces from 10 Latin music genres was developed. This database is called the Latin Music Dataset. The experimental results show that the proposed approach allows correctly classifying 66.76% of the music pieces. This represents an improvement of 9.33% over the previous approach using the best individual classifier that considers only the features extracted from the beginning of the songs. By applying the feature selection process we achieved an improvement of 7.43%.

Keywords: Automatic Music Genre Classification, Ensemble of Classifiers, Feature Selection, Pattern Recognition of Audio Signals.

Capítulo 1

Introdução

Com a rápida expansão da Internet uma imensa massa de dados oriundos de diferentes fontes tem se tornado disponível *on-line*. Um estudo da Universidade de Berkeley (LYMAN; VARIAN, 2003) mostra que em 2002 haviam cinco milhões de terabytes de novas informações criadas em documentos impressos, filmes, mídias ópticas e magnéticas. Estima-se que a WWW sozinha contenha cerca de 170 terabytes de informação.

Porém, toda esta informação não segue um padrão de apresentação e não está disponível de maneira estruturada, o que torna muito difícil fazer uso adequado da mesma. Devido a isto, tarefas como busca, recuperação, indexação, extração e sumarização automática dessas informações se tornaram problemas importantes sobre os quais muitas pesquisas têm sido realizadas. Neste contexto uma área de pesquisa que tem crescido nos últimos anos é a de recuperação de informações multimídia que visa criar ferramentas capazes de organizar e gerenciar essa grande quantidade de informações (FINGERHUT, 1999) (PAMPALK; RAUBER; MERKL, 2002). No momento, a maior parte das informações sobre dados multimídia são organizadas e classificadas baseadas em meta-informações textuais que são associadas ao seu conteúdo, como é o caso dos rótulos ID3 incorporados nos arquivos de áudio no formato MP3. Apesar destas informações serem relevantes para as tarefas de indexação, busca e recuperação, elas dependem da intervenção humana para gerá-las e posteriormente associá-las aos arquivos multimídia.

A música digital é um dos mais importantes tipos de dados distribuídos na Internet. Existem muitos estudos e métodos a respeito da análise de conteúdo de áudio usando diferentes características e métodos (PAMPALK; RAUBER; MERKL, 2002), (GUO; LI, 2003), (ZHANG; KUO, 2001), (LI; OGIHARA; LI, 2003), (AUCOUTURIER; PACHET, 2003). Um componente fundamental para um sistema de recuperação de informações de áudio baseado em conteúdo é um módulo de classificação automática de gêneros musicais (LI; OGIHARA, 2005), visto que os gêneros musicais têm sido utilizados para classificar e caracterizar

músicas digitais e para organizar grandes coleções disponíveis na Web (LI; OGIHARA; LI, 2003), (AUCOUTURIER; PACHET, 2003), (TZANETAKIS; COOK, 2002), (SHAO; XU; KANKANHALLI, 2003).

1.1 Definição do Problema

Os gêneros musicais são rótulos categóricos criados por especialistas humanos, assim como por amadores, para determinar ou designar estilos de música. Esses rótulos estão relacionados com a instrumentação/orquestração utilizada, estrutura rítmica e conteúdo harmônico da música.

Inicialmente pode-se pensar que as informações existentes no campo *Genre* (Gênero) do cabeçalho ID3 das músicas no formato MP3 sejam suficientes. Porém a primeira versão do formato ID3 conhecido como ID3v1, possui uma lista fixa de 80 gêneros. Como este campo é representado por um byte, existe a possibilidade de incluir nesta lista outros gêneros, e é justamente isto que acontece na maioria dos programas que permitem editar o cabeçalho ID3 dos arquivos MP3. O problema é que não existe um padrão para listar os gêneros. Cada fabricante utiliza uma seqüência diferente para listar os gêneros acima do 80, fazendo com que em um determinado programa uma música que está classificada como, por exemplo, *Tango* apareça como sendo *Trash Metal* quando utilizada/tocada/executada em outro programa. Desta forma, as informações cadastradas no campo gênero dos rótulos ID3 normalmente não são confiáveis. Mais informações sobre o formato ID3 podem ser encontradas no Anexo A.

Porém, a questão do ID3 não é o único problema encontrado na definição de gêneros musicais, visto que um gênero musical é um conceito relativamente subjetivo. Mesmo a indústria musical muitas vezes é contraditória ao atribuir gêneros musicais para as músicas. Adicionalmente, a atribuição de gêneros musicais tem sido feita para álbuns, e esta atribuição não é obrigatoriamente aplicável às faixas do álbum (AUCOUTURIER; PACHET, 2003). Desta forma, a classificação automática de gêneros musicais pode auxiliar ou substituir o usuário humano neste processo, assim como prover um componente importante para um sistema de recuperação de informações para músicas, podendo também tornar menos subjetivo este processo de atribuição.

Estudos sobre o comportamento de usuários de Sistemas de Recuperação de Informação Multimídia, também conhecidos como sistemas MIR (*Multimidia Information Retrieval*) (DOWNIE; CUNNINGHAM, 2002), indicam que o gênero musical é o segundo item mais fornecido nas *strings* (*queries*) de busca. O primeiro item é relacionado às informações biográficas da música. Nos trabalhos de (VIGNOLI, 2004) e (LEE; DOWNIE,

2004) foi verificado que o gênero musical é freqüentemente visto pelos usuários como um importante método de organização de coleções musicais e de recuperação de músicas de grandes coleções.

1.2 Objetivo

O principal objetivo deste trabalho é o desenvolvimento de um sistema para a classificação automática de gêneros musicais baseado no conteúdo do sinal de áudio das músicas. Para realizar esta tarefa são utilizados algoritmos clássicos de aprendizagem supervisionada.

Visando reduzir o esforço computacional necessário para a extração das características do sinal de áudio das peças musicais, são avaliados métodos de seleção de atributos. Observa-se a relevância das características inicialmente escolhidas na tarefa de classificação de gêneros musicais, ou se é possível utilizar apenas um subconjunto das mesmas, desta forma reduzindo o número de características a serem computadas a partir do sinal de áudio. Neste sentido, pretende-se verificar o desempenho de um método de seleção de atributos baseado em algoritmos genéticos.

Levando-se em conta que a precisão é um aspecto importante do sistema, são avaliados métodos de combinação de classificadores visando aumentar o desempenho. Mais especificamente, são utilizados os métodos baseados na decomposição do sinal de áudio (*Time Decomposition*), na decomposição do espaço do problema (*Space Decomposition*) e nos escores de confiança fornecidos na saída de cada classificador.

1.3 Desafio

Os principais desafios a serem superados para o desenvolvimento deste trabalho são:

- O desenvolvimento de uma nova base de dados para a tarefa contendo pelo menos 3.000 músicas de 10 gêneros distintos, sendo pelo menos 300 músicas de cada gênero;
- A implementação de um *framework* confiável para o processo de criação e manutenção desta base de dados, tendo em vista o esforço humano necessário para o processo de rotulação da base;
- A definição de uma forma adequada para realizar a extração de características das peças musicais de forma que não seja necessário realizar esta etapa para cada um

dos experimentos a serem realizados;

- O treinamento e execução dos diversos classificadores, o uso de técnicas de combinação de classificadores, e a avaliação comparativa de todos estes resultados.

1.4 Hipótese

A primeira hipótese deste trabalho é que o uso de características extraídas de diferentes partes da música, a utilização de classificadores e posteriormente a combinação da saída dos mesmos leva a uma classificação mais eficiente dos gêneros musicais das músicas em relação a outras abordagens que utilizam somente características extraídas de um trecho da música (normalmente o início) ou características extraídas das músicas completas. A segunda hipótese deste trabalho é que o vetor de características inicial pode ser otimizado através do uso de algoritmos de seleção de características melhorando a taxa de classificação dos algoritmos de classificação. A terceira hipótese deste trabalho é que o uso conjunto das técnicas de combinação de classificadores e seleção de atributos deve melhorar a taxa de classificação dos algoritmos de classificação em relação às abordagens anteriores.

1.5 Organização do Documento

Esse trabalho está organizado em cinco capítulos. O Capítulo 2 apresenta uma revisão dos principais conceitos utilizados no método proposto para a classificação automática de gêneros musicais. O Capítulo 3, por sua vez, descreve a abordagem proposta para solucionar o problema da classificação automática de gêneros musicais utilizando seleção de atributos e combinação de classificadores. O Capítulo 4 apresenta os experimentos realizados, bem como os resultados obtidos. Por último, as conclusões e direções futuras do trabalho são apresentadas no Capítulo 5.

Capítulo 2

Reconhecimento de Padrões e Classificação Automática de Gêneros Musicais

Neste capítulo são apresentados os principais conceitos relacionados ao desenvolvimento deste trabalho assim como uma análise crítica dos principais trabalhos relacionados.

2.1 Reconhecimento de Padrões

A tarefa de RP (*Reconhecimento de Padrões*) tem como objetivo atribuir rótulos para objetos. Os objetos, por sua vez, são descritos por um conjunto de medidas denominadas características ou atributos. Porque RP se defronta com os desafios encontrados em problemas da vida real, apesar das décadas de pesquisa produtiva, teorias modernas e elegantes ainda co-existem com idéias ad-hoc, intuição e palpites (*guessing*). Isto é refletido pela variedade de métodos e técnicas disponíveis aos pesquisadores.

A Figura 2.1 ilustra as tarefas básicas e os estágios da tarefa de reconhecimento de padrões. Suponha que um usuário hipotético nos apresenta um problema e um conjunto de dados. Nossa tarefa é clarificar o problema, traduzí-lo para a terminologia de reconhecimento de padrões, resolvê-lo e comunicar a solução de volta para o usuário.

Existem dois tipos de problemas de reconhecimento de padrões: supervisionados e não-supervisionados. Na categoria de problemas não-supervisionados (também conhecidos como aprendizado não-supervisionado), o problema é descobrir a estrutura do conjunto de dados se houver alguma. Isto usualmente significa que o usuário quer conhecer se existem grupos nos dados, e quais características fazem os objetos similares em um grupo e dissimilares dos demais grupos.

Na categoria de problemas supervisionados, cada objeto no conjunto de dados possui uma classe previamente rotulada. A tarefa é treinar um classificador para rotular novos objetos de acordo com estas classes.

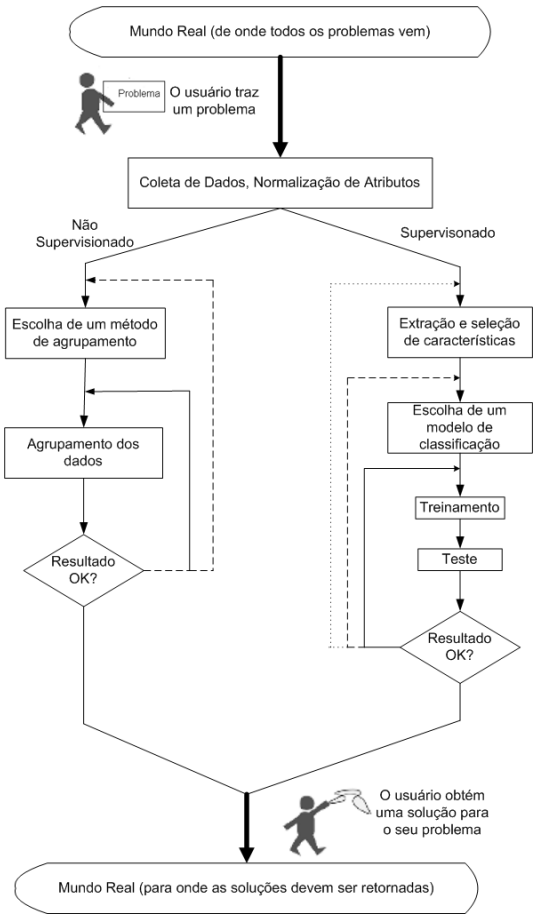

Figura 2.1: Visão Geral das Tarefas de Reconhecimento de Padrões (KUNCHEVA, 2004)

2.2 Combinação de Classificadores

As principais razões para a combinação de classificadores são eficiência e precisão (KITTLER et al., 1998). Nesse trabalho Kittler et al. apresenta dois cenários para a combinação de classificadores. No primeiro cenário, todos os classificadores usam a mesma representação do padrão de entrada. Apesar de cada classificador utilizar o mesmo vetor de características, cada classificador vai lidar com ele de formas diferentes. Essa idéia é ilustrada com dois exemplos: no primeiro poderia ser utilizado um conjunto de classificadores baseados nos k vizinhos mais próximos k-NN (*k Vizinhos Mais Próximos*) onde cada classificador tem um número diferente para o valor de k. O segundo exemplo seria utilizar um conjunto de redes neurais, cada uma treinada com uma função de treinamento diferente ou mesmo com funções similares, porém com parâmetros diferentes. No segundo cenário, cada classificador utiliza sua própria representação do padrão de entrada.

No contexto do segundo cenário, uma estratégia possível para resolver problemas de classificação multi-classe é o uso de técnicas de decomposição do espaço do problema (*Problem-Space Decomposition*). O principal motivo para aplicar qualquer método de

decomposição do problema é que a classificação multi-classe é intrinsicamente mais difícil do que a classificação binária, pois o algoritmo de classificação tem que construir um grande número de superfícies de separação, enquanto que os classificadores binários têm que determinar apenas uma função adequada de decisão (DIETTERICH, 2000).

2.2.1 Um-Contra-Todos (One-Against-All)

Dado um problema de reconhecimento de padrões de *nc*-classes, a estratégia OAA (*Um-Contra-Todos (One-Against-All)*) consiste em criar um conjunto de *nc* classificadores binários, um para cada classe. Cada classificador é treinado através de um processo de re-rotular o mesmo conjunto de treinamento, de forma a distingüir entre uma classe e o seu complemento no espaço do problema. Por exemplo, o classificador para a classe C_i é treinado utilizando os elementos da classe C_i como exemplos positivos e o restante do conjunto de treinamento como exemplos negativos, produzindo um classificador especializado para a classe C_i.

Para um exemplo desconhecido, representando por um vetor de características x, dadas as *nc* classificações individuais, e considerando que cada classificador individual atribui a x uma probabilidade p (ou um *score* de confiança) que está diretamente relacionado a conformidade deste exemplo com sua classe, a classe final $\hat{C}$ atribuída a x vai ser dada por:

$$\hat{C} = \arg\max_{1 \leq i \leq nc} p(C_i|x) \tag{2.1}$$

onde $p(C_i|x)$ é a probabilidade a *posteriori* da classe C_i dado um vetor de características x e $\hat{C}$ é a classe vencedora, isso é, aquela que fornece a maior probabilidade a *posteriori*. A Figura 2.2 ilustra esse processo.

2.2.2 Round Robin

No trabalho de Fürnkranz (2002), a técnica de decomposição de problemas intitulada RR (*Round Robin*) é utilizada para a combinação de classificadores, de forma a permitir que classificadores binários lidem com problemas multi-classe. O método de *Round Robin* converte um problema de *nc*-classes em uma série de problemas binários, criando um conjunto de $l = \frac{nc(nc-1)}{2}$ classificadores, um para cada par de classes.

Novos exemplos são classificados após serem apresentados ao conjunto de l classificadores binários. Nesse caso, quando um novo exemplo e é apresentado para cada um dos l classificadores binários, uma classe é atribuída a e. As l classes atribuídas vão ser

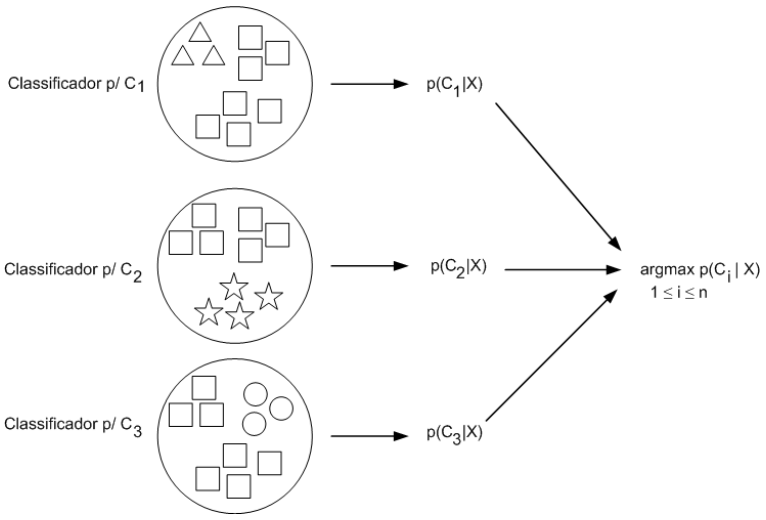

Figura 2.2: Ilustração da estratégia *One-Against-All* para um problema de três classes

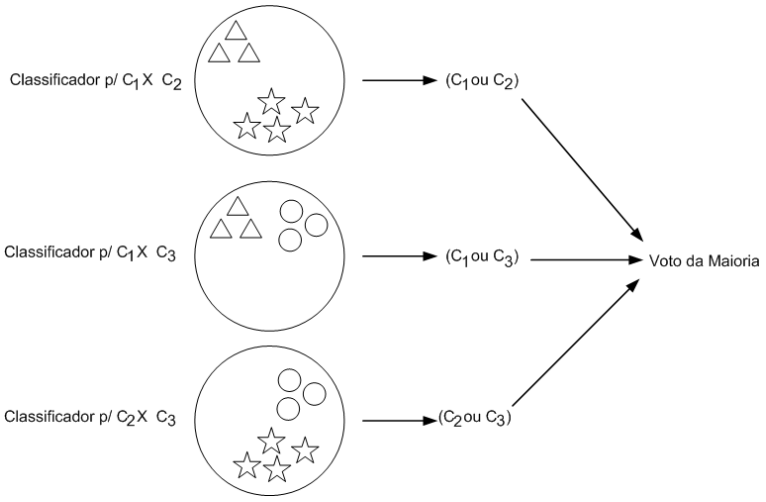

Figura 2.3: Ilustração da estratégia *Round Robin* para um problema de três classes

combinadas e o resultado final é obtido através do voto da maioria, ou seja, a classe final vai ser definida como sendo aquela que teve o maior número de ocorrências nas respostas dos classificadores.

Ao contrário da estratégia de Um-Contra-Todos (*One-Against-All*), neste caso quando um classificador binário é construído, vamos dizer para as classes C_i e C_j, apenas os exemplos destas duas classes são utilizadas, e o resto do conjunto de treinamento é ignorado. De acordo com Fürnkranz (2002), esta abordagem leva a uma superfície de decisão mais simples sobre os limites das duas classes. A Figura 2.3 ilustra essa abordagem.

2.2.3 Combinação Baseada na Saída dos Classificadores (Output Level Ensemble)

Uma outra técnica de combinação de classificadores é baseada nos escores de confiança fornecidos por cada classificador para cada classe (KITTLER et al., 1998). O uso

desta técnica requer diferentes vetores de características para o mesmo objeto, e formalmente pode ser definida como:

Considerando uma seqüência de M vetores de características extraídos de um objeto como sendo

$$X_t =< \bar{x}_D(1), \bar{x}_D(2), \ldots, \bar{x}_D(M) > \tag{2.2}$$

no qual cada $\bar{x}_D(m)$ é um vetor de características relacionado a uma representação m, onde $m = 1, 2, \ldots, M$. De maneira similar é possível definir uma seqüência de M classificadores como sendo

$$C =< c(1), c(2), \ldots, c(M) > \tag{2.3}$$

Sem perda de generalidade é assumido que este conjunto de classificadores probabilísticos é homogêneo, cuja saída de cada classificador é $P(r|X_t)$, onde $\sum_{r \in \mathcal{R}} P(r|X_t) = 1$. Sendo r um rótulo possível dentro todo o conjunto de rótulos possíveis R. A relação entre X_t e C direta, ou seja, é uma relação de um-para-um, e o vetor de características m da seqüência de vetores X_t é classificado pelo classificador $c(m)$ de C.

De forma a encontrar a melhor combinação de classificadores, ou seja, o conjunto de classificadores mais diverso que produz uma boa generalização, é utilizada uma única função objetivo baseada na maximização da taxa de reconhecimento da combinação dos classificadores.

As regras de combinação utilizadas neste trabalho são:

- MAJ (*Voto da Maioria*): Nesta regra de decisão, apenas os rótulos das classes são levadas em consideração e aquela com o maior número de votos ganha[1]:

$$\hat{C} = maxcount \left[\underset{\substack{r \in \mathcal{R} \\ m \in [1,\ldots,M]}}{\arg\max} \; P_m(r|x_m^D) \right] \tag{2.4}$$

- MAX (*Máximo*): Nesta regra de decisão, a classe com o maior *score* de confiança é escolhida. Formalmente essa regra é definida por:

$$\hat{C} = \underset{\substack{r \in \mathcal{R} \\ m \in [1,\ldots,M]}}{\arg\max} \; P_m(r|x_m^D) \tag{2.5}$$

[1] A função *maxcount* retorna o valor mais freqüente de um multiset.

- SUM (*Soma*): A regra do somatório é baseada nos escores de confiança produzidos na saída dos classificadores para todas as classes. Os escores de confiança serão somados para cada classe e a classe com o maior valor é escolhida. Formalmente essa regra é definida por:

$$\hat{C} = \arg\max_{r \in \mathcal{R}} \sum_{m=1}^{M} P_m(r|x_m^D) \tag{2.6}$$

- WS (*Soma Ponderada*): Ao invés de utilizar apenas a regra do somatório simples, é possível adicionar pesos à saída de cada classificador. Neste caso os escores de confiança fornecidos por cada classificador são multiplicados por constantes $(\alpha, \beta, \ldots, \mu)$:

$$\hat{C} = \arg\max_{r \in \mathcal{R}} \alpha P_1(r|x_m^D) + \ldots + \mu P_M(r|x_m^D) \tag{2.7}$$

- PROD (*Produto*): A regra da multiplicação é baseada nos escores de confiança produzidos na saída dos classificadores para todas as classes. Os escores de confiança para cada classe são multiplicados e a classe com o maior valor é escolhida:

$$\hat{C} = \arg\max_{r \in \mathcal{R}} \prod_{m=1}^{M} P_m(r|x_m^D) \tag{2.8}$$

- WP (*Produto Ponderado*): Essa regra utiliza a mesma idéia dos pesos utilizada pela regra WS. Porém, ao invés de multiplicar os pesos a saída de cada classificador, os escores de confiança são elevados ao valor dos pesos.

$$\hat{C} = \arg\max_{r \in \mathcal{R}} P_1(r|x_m^D)^{\alpha} \times \ldots \times P_M(r|x_m^D)^{\mu} \tag{2.9}$$

A Figura 2.4 apresenta um exemplo com um problema de três classes, utilizando três classificadores, em que o resultado final depende da regra de combinação utilizada.

2.3 Seleção de Atributos

Com o intuito de melhorar a taxa de reconhecimento dos algoritmos de classificação é possível efetuar uma etapa de seleção de atributos. Esta etapa consiste em escolher um subconjunto de nf características do conjunto original de D características ($nf \leq D$), de forma que o espaço de busca seja reduzido de acordo com algum critério de otimização

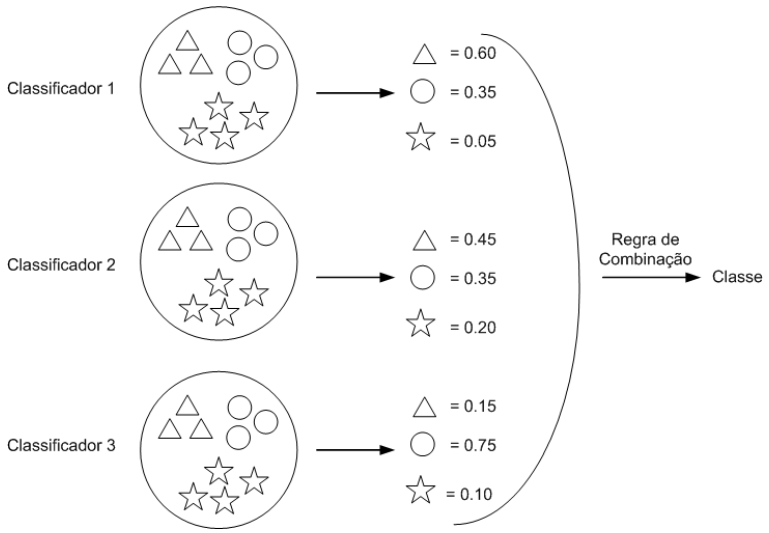

Figura 2.4: Ilustração da estratégia de combinação baseada na probabilidade a posteriori dos classificadores

(BLUM; LANGLEY, 1997). O papel da seleção de atributos nas tarefas de reconhecimento de padrões é:

- Reduzir a dimensionalidade do espaço de busca;
- Acelerar o algoritmo de aprendizado;
- Melhorar o acerto preditivo de um algoritmo de classificação;
- Melhorar a compreensibilidade dos resultados de aprendizado.

De forma geral, a seleção de atributos é um problema de otimização de acordo com algum critério de avaliação. Um método típico de seleção de atributos consiste de quatro etapas básicas (como apresentado na Figura 2.5: geração de subconjuntos, avaliação dos subconjuntos, critério de parada, e validação dos resultados (DASH; LIU, 1997)). Na primeira etapa (geração) o algoritmo tem como entrada o conjunto original de características a partir do qual é gerado um subconjunto de características. Na segunda etapa este subconjunto é avaliado e recebe um valor referente a sua qualidade. A terceira etapa é um processo de decisão, se algum critério de parada foi atingido o procedimento termina e este subconjunto é submetido à validação na quarta etapa, caso contrário, volta-se para a primeira etapa. Na quarta etapa o subconjunto gerado é utilizado nos experimentos para verificar se traz benefícios em relação ao uso do vetor de características.

A geração de subconjuntos é um procedimento de busca. Basicamente, ela gera subconjuntos de características para avaliação. Seja D o número total de características

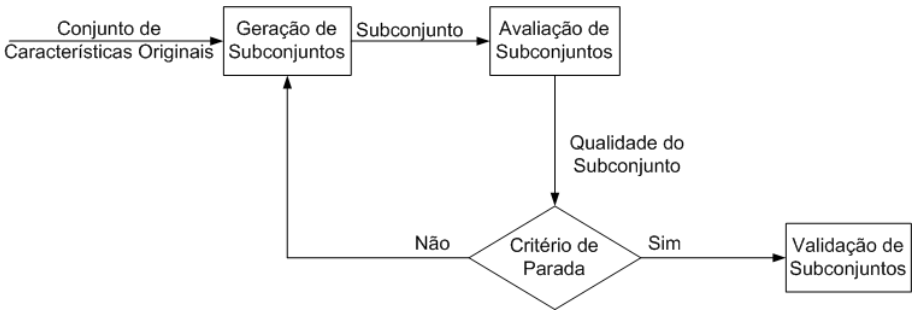

Figura 2.5: Etapas Básicas de um método de seleção de características. (DASH; LIU, 1997)

no conjunto original, então o número total de subconjuntos candidatos é 2^D, o que faz com que o método de busca exaustiva no espaço de características seja inviável mesmo com um número moderado de D.

Os métodos de seleção de atributos podem ser classificados em três grupos de acordo com como as características são utilizadas e avaliadas (MOLINA; BELANCHE; NEBOT, 2002):

- Incorporado (*Embbedded*): O algoritmo possui um mecanismo próprio para seleção de atributos. É o caso, por exemplo, dos algoritmos de árvores de decisão;
- Filtro (*Filter*): O processo de seleção de atributos acontece antes de qualquer algoritmo de reconhecimento de padrões ser utilizado. De uma forma geral, diz-se que o processo de seleção de atributos ocorre na etapa de pré-processamento.
- Envelope (*Wrapper*): Nessa abordagem, o algoritmo de aprendizado de máquina é utilizado como uma sub-rotina. O algoritmo específico é utilizado para avaliar as soluções geradas.

(Na figura em negrito) A figura 2.6 ilustra os métodos de seleção de atributos utilizando as abordagens de filtro e de envelope. Na figura 2.6, blocos em ênfase, é possível verificar o mecanismo de seleção de atributos, no caso da abordagem filtro, este encontra um subconjunto de características ótimas sem utilizar o algoritmo de aprendizagem. Já no caso da abordagem de envelope é possível ver como a geração dos subconjuntos é um processo iterativo, onde dada as características originais, vai ser gerado um subconjunto de características, este subconjunto vai ser apresentado para o algoritmo de aprendizagem de máquina que vai retornar uma avaliação do subconjunto. Este processo é repetido até que seja encontrado um subconjunto ótimo.

Uma outra opção para gerar subjconjuntos é uso de AGs (*Algoritmos Genéticos*), pois eles oferecem uma busca aleatória guiada (*random guided search*) no espaço de todos os possíveis subconjuntos. A forma mais conveniente de representar um subconjunto de

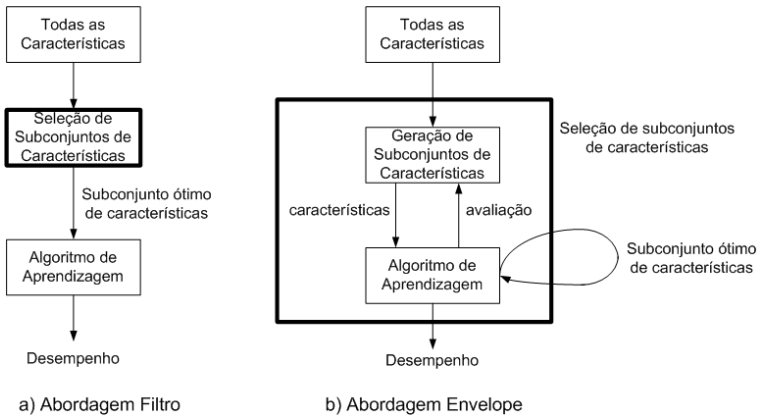

Figura 2.6: Ilustração dos métodos de seleção de atributos. (YANG; HONAVAR, 1998)

características é utilizando um vetor binário de tamanho D. A i-ésima posição do vetor é 1 se o i-ésimo atributo estiver incluído no subconjunto e 0 caso contrário. Um AG opera sobre um conjunto de M vetores binários, denominados cromossomos. A população de cromossomos é evoluída através do uso de operadores genéticos denominados mutação e crossover.

Seja $X = x_1, \ldots, x_{10}$ o conjunto de características. Uma população com 6 cromossomos (indivíduos), $S_1, \ldots, S_6$ é apresentada abaixo (KUNCHEVA, 2004):

$$
\begin{array}{lcccccccccc}
S_1 & 0 & 1 & 0 & 1 & 0 & 0 & 0 & 1 & 0 & 0 & \{x_2, x_4, x_8\} \\
S_2 & 1 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & \{x_1, x_2, x_3, x_{10}\} \\
S_3 & 0 & 0 & 1 & 1 & 1 & 1 & 0 & 0 & 0 & 1 & \{x_3, x_4, x_5, x_6\} \\
S_4 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 1 & \{x_8, x_{10}\} \\
S_5 & 0 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & \{x_2, x_3\} \\
S_6 & 1 & 1 & 0 & 1 & 1 & 1 & 0 & 0 & 1 & 1 & \{x_1, x_2, x_4, x_5, x_6, x_9, x_{10}\}
\end{array}
$$

Para avaliar a adequabilidade (*Fitness*) dos indivíduos, um classificador vai ser treinado utilizando as características selecionadas pelo indivíduo e seu desempenho representa o valor de adequabilidade (abordagem do envelope (*wrapper*)).

2.4 Classificação Automática de Gêneros Musicais

Nesta seção são apresentados além do estado da arte, os principais trabalhos da área de classificação automática de gêneros musicais.

2.4.1 Sem combinação de classificadores

A idéia de classificação automática de gêneros musicais como uma tarefa de reconhecimento de padrões de sinais de músicas foi apresentada no trabalho de Tzanetakis e Cook (2002). Neste trabalho, foi proposto um conjunto abrangente de características para representar um sinal de áudio. As características foram utilizadas para construir três tipos de classificadores: classificador Gaussiano, modelos de mistura Gaussiana e k-NN. O conjunto de características proposto é composto de características relacionadas ao espectro sonoro, ao padrão rítmico (*beat-related*) e à altura da nota (*pitch-related*). Os experimentos foram avaliados numa base de dados contendo 1.000 músicas de 10 gêneros distintos, sendo 100 músicas de cada gênero. Os gêneros utilizados foram: *Blues*, *Classical*, *Country*, *Disco*, *Hiphop*, *Jazz*, *Metal*, *Pop*, *Reggae* e *Rock*. O acerto obtido inicialmente nesta base foi de cerca de 60% utilizando o mecanismo de validação-cruzada fator 10 cem vezes. É importante observar que os experimentos foram avaliados utilizando apenas os primeiros 30 segundos de cada música. Outro aspecto interessante deste trabalho é que o conjunto de características utilizado está disponível através do Framework Marsyas[2] (TZANETAKIS; COOK, 1999), um software livre para o desenvolvimento e avaliação de aplicações voltadas à computação musical.

Tzanetakis e Cook (2002) motivaram a pesquisa e desenvolvimento de novas abordagens para a tarefa de classificação automática de gêneros musicais utilizando técnicas de aprendizado de máquina e processamento digital de sinais.

No trabalho de Kosina (2002) foi desenvolvido o MUGRAT[3] – um sistema protótipo para a classificação automática de gêneros musicais que utiliza um subconjunto de características das propostas por Tzanetakis & Cook. A principal diferença do MUGRAT para o Marsyas, é que ele não utiliza as características relacionadas ao *pitch* nem as relacionadas aos cinco primeiros MFCC (*Coeficientes Cepstrais de Freqüência-Mel*). A avaliação do MUGRAT foi efetuada numa base de dados contendo 189 músicas de três gêneros: *Metal* (63), *Dance* (65) e *Classical* (61). Os vetores de características foram extraídos de amostras aleatórias com três segundos de duração. A melhor classificação (88.35%) foi obtida utilizando o classificador k-NN com o valor de $k = 3$ com o método de avaliação de validação-cuzada estratificada fator 10. Um aspecto interessante constatado no trabalho é que ao construir a base de dados a autora percebeu que a mesma música no formato MP3 obtida de fontes diferentes possuía uma informação diferente no campo gênero (*Genre*) do rótulo ID3. Este fato foi utilizado para ilustrar que a classificação humana de gêneros

[2]Disponível em: http://marsyas.sourceforge.net/
[3]Disponível em: http://kyrah.net/mugrat/

musicais realmente é nebulosa, mas como mostrado na seção introdução, este não é o único problema associado aos rótulos ID3.

Li, Ogihara e Li (2003) realizaram um estudo comparativo para a classificação automática de gêneros musicais baseada em conteúdo entre o conjunto de características propostas por Tzanetakis & Cook e um novo conjunto de características extraídos utilizando DWCH (*Histogramas de Coeficientes fornecidos pela Daubechies Wavelet*). Eles também desejam verificar se outros métodos estatísticos como LDA (*Análise discriminante linear*) e SVM (*Máquinas de suporte vetorial*) teriam um melhor desempenho do que os demais classificadores utilizados anteriormente. Os experimentos foram realizados em duas bases de dados: a primeira (Base A) é a mesma utilizada nos experimentos de Tzanetakis & Cook e a segunda (Base B) contém 756 músicas de cinco gêneros (*Ambient* (109), *Classical* (164), *Fusion* (136), *Jazz* (251) e *Rock* (96)). Um aspecto importante desta segunda base de dados é que as características foram extraídas do segmento composto pelo segundo 31 ao segundo 61, ao invés dos primeiros trinta segundos (como acontece na base do Tzanetakis & Cook). As conclusões dos experimentos realizados neste trabalho mostram que o melhor resultado foi obtido com o classificador SVM que melhorou o acerto obtido na base A para cerca de 72% com o mesmo conjunto de características, e para cerca de 78% no melhor caso com as características da DWCH (em ambos os casos utilizando o classificador SVM). Na base B a taxa de acerto obtida foi de 74% (utilizando DWCH) e 71% (utilizando as características do Tzanetakis & Cook). Outro aspecto importante deste trabalho é que eles avaliaram diferentes estratégias de decomposição que são necessárias por classificadores que não lidam naturalmente com problemas multi-classe. Eles avaliaram o classificador SVM utilizando as estratégias de Um-Contra-Todos (OAA), Round Robin(RR)(que eles chamam de *Pairwise Comparison*) e funções objetivas multi-classe. Os melhores resultados foram alcançados com a estratégia OAA com as características da DWCH. A diferença entre a taxa de classificação obtida pelo conjunto de características baseado em DWCH em relação às características do Tzanetakis & Cook foi de 2% (utilizando o k-NN) a 7% utilizando SVM com OAA para a base A e de 2% (utilizando o k-NN) a 4% utilizando SVM com OAA para a base B.

No trabalho de Li e Ogihara (2005) foi investigado o uso de uma taxonomia hierárquica para a classificação de gêneros musicais. Esta taxonomia identifica as relações de dependência de diferentes gêneros e fornece valiosas fontes de informação para a classificação de gêneros. Os experimentos foram realizados com as mesmas bases utilizadas anteriormente pelo grupo (LI; OGIHARA; LI, 2003) e a taxa de classificação aumentou em 0.7% para a base A e 3% para a base B.

Em (SILLA JR.; KAESTNER; KOERICH, 2005) foram avaliados métodos de *Bagging* e

Boosting aliados aos classificadores J48 (*Árvores de Decisão*), NB (*Naive Bayes*) e 3-NN. Os experimentos foram realizados utilizando a base do trabalho de Tzanetakis e Cook (2002). O uso das técnicas de meta-aprendizagem aumentaram a taxa de classificação correta do J48 em todos os casos. Para o NB os métodos de meta-aprendizagem se mostraram ineficientes, enquanto que para o 3-NN apenas o método de *Bagging* forneceu melhores resultados.

Um trabalho relacionado com a tarefa de classificação automática de gêneros musicais, porém com outro foco, é o realizado por Hu et al. (2005) onde são utilizados *reviews* de músicas e técnicas de mineração de textos para realizar a classificação automática dos gêneros.

2.4.2 Com combinação de classificadores

A idéia de decomposição e combinação de classificadores foi utilizada para a classificação automática de gêneros musicais no trabalho de Grimaldi, Cunningham e Kokaram (2003b, 2003a). Nestes trabalhos foram realizados experimentos utilizando diferentes estratégias de combinação de classificadores e seleção de atributos. Eles avaliaram o desempenho de OAA, RR e RSM (*Método de Busca em Subespaços Aleatórios*) (HO, 1998) com alguns algoritmos para ranqueamento de características para seleção de atributos, conhecidas como PCA (*Análise de Componentes Principais*), IG (*Ganho de Informação*) e GR (*Razão de Ganho*). Os experimentos foram realizados numa base contendo 200 músicas de cinco gêneros (Jazz, Classical, Rock, Heavy Metal e Techno). Para efetuar a validação foi utilizado o método de validação cruzada fator 5. Todos os experimentos foram avaliados utilizando apenas o classificador k-NN. Para extrair as características foi utilizada a DWPT (*Transformada Wavelet Discreta*) aplicada ao sinal da música inteira.

No trabalho de Costa, Valle Jr. e Koerich (2004) foi proposto um novo método para a classificação automática de gêneros musicais, baseado na extração de características de três segmentos do sinal de áudio. As características foram extraídas do início, meio e fim da música. Para cada segmento foi treinado um classificador componente e a decisão final era obtida através do voto da maioria de cada uma das partes. Os classificadores utilizados foram MLP (*Rede Neural do tipo Multi-Layer Perceptron*) e k-NN. As características foram extraídas utilizando o software MUGRAT. Os experimentos foram realizados em uma base contendo 414 músicas de dois gêneros (*Rock* e *Classical*). A base foi particionada em três conjuntos: treinamento com 208 músicas, validação com 82 músicas e teste com 122 músicas. A conclusão obtida no trabalho foi que o método de combinação proposto não melhorava o desempenho além da classificação individual dos

segmentos isolados.

Uma continuação do trabalho de Costa, Valle Jr. e Koerich (2004) foi apresentada por Koerich e Poitevin (2005) onde para realizar a combinação dos classificadores foram utilizadas outras regras de combinação além do voto da maioria. As regras eram baseadas nas probabilidades individuais de cada classe fornecida na saída dos classificadores. As regras utilizadas foram MAX, SUM, WS, PROD e WP. A base utilizada foi a mesma do experimento anterior. Uma alteração é que neste trabalho os autores utilizaram apenas redes neurais MLP para fazer a classificação. Os resultados obtidos mostraram uma melhora na taxa de acerto em relação aos segmentos individuais utilizando dois segmentos e as regras de soma e produto ponderados.

No trabalho de Meng, Ahrendt e Larsen (2005) são utilizadas características baseadas em três escalas de tempo: as características de tempo curto são computadas utilizando janelas de análise de tamanho 30 ms, o significado perceptual deste tipo de característica está relacionado ao timbre (freqüência instantânea); as características de tempo médio são computadas utilizando janelas de análise de tamanho 740 ms, e estão relacionadas à modulação (instrumentação); as características de tempo longo são computadas utilizando janelas de análise de tamanho 9.62s e estão relacionadas à batida, ao padrão rítmico e inflexão vocal, etc. Para realizar os experimentos foram considerados dois classificadores: LNN (*Rede neural simples de uma camada*) e um classificador Gaussiano com uma matriz completa de covariância. Os experimentos foram realizados em duas bases de dados, mas o propósito destes era verificar o desempenho relativo das características ao invés de verificar o erro no conjunto de dados. A primeira base de dados utilizada contém 100 músicas, distribuídas igualmente em cinco gêneros (*Classical*, *Rock*, *Jazz*, *Pop* e *Techno*). A segunda base consiste de 354 músicas de 30 segundos extraídas do "Amazon.com Free-Downloads" e possuem 6 gêneros (*Classical*, *Country*, *Jazz*, *Rap*, *Rock* e *Techno*). Foram realizados diversos experimentos e os melhores resultados computacionais obtidos no conjunto de teste foram de 5% em relação a 1ª base de dados utilizando a combinação de características de tempo médio e longo.

No trabalho de Yaslan e Cataltepe (2006) foram utilizados os seguintes classificadores: Fisher (Classificador de Fisher); LDC (*Linear classifier assuming normal densities with equal covariance matrices*); QDC (*Quadratic classifier assuming normal densities*); UDC (*Quadratic classifier assuming normal uncorrelated densities*); NB (Classificador Naïve Bayes); PDC (*Parzen Density Based Classifier*); k-NN (Vizinhos mais próximos com o valor ótimo de k computado utilizando o metodo de validação cruzada com *leave-one-out*); 1-NN (1 vizinho mais próximo), 3-NN (3 vizinhos mais próximos); 5-NN (5 vizinhos mais próximos). A base utilizada foi a GTZAN (*Base de dados desenvolvida no*

trabalho de Tzanetakis e Cook (2002)) e o processo de extração de características foi efetuado com o MARSYAS. A principal diferença desse trabalho em relação aos anteriores, é que foram avaliadas as características de acordo com os grupos a que elas pertecem para cada um dos classificadores listados. Além disso foram utilizados métodos de FFS (*Seleção de características com busca para frente*) e BFS (*Seleção de características com busca para trás*) para tentar encontrar um melhor subconjunto de características que aumentasse o desempenho dos classificadores. Os resultados obtidos foram positivos e os autores ainda propuseram o uso de um *ensemble* (que deveria ter sido classificado como *Stacking*) combinando a saída dos classificadores que apresentarem os melhores resultados. Essa técnica de combinação também apresentou resultados positivos.

2.4.3 Com seleção de características

O uso de métodos de seleção de características para a classificação automática de gêneros musicais foi recentemente avaliado nos trabalhos de (FIEBRINK; FUJINAGA, 2006) e (YASLAN; CATALTEPE, 2006). No trabalho de Fiebrink e Fujinaga (2006) foram realizados experimentos utilizando métodos de FFS e PCA em conjunto com o classificador k-NN para classificar a base Magnatune[4](4.476 amostras de 24 gêneros), com 74 características que foram extraídas utilizando o JAudio (MCENNIS et al., 2005). As conclusões obtidas neste trabalho foram que considerando o desempenho dos sistemas utilizando PCA os resultados obtidos foram similares ao uso do método de FFS porém com um tempo computacional bem reduzido.

2.4.4 Recursos e Ferramentas

Alguns trabalhos recentes apresentam uma preocupação com o desenvolvimento de ferramentas que possam trabalhar diretamente com a classificação automática de gêneros musicais e que isto possa ser feito de forma escalável (devido à grande quantidade de recursos computacionais necessários). A versão mais nova do Marsyas desenvolvida por Bray e Tzanetakis (2005) foi projetada para trabalhar com diferentes computadores de forma a distribuir a carga computacional. Visando o desenvolvimento de um *framework* comum para o desenvolvimento de extração de características a partir de sinais de áudio McEnnis et al. (2005) desenvolveram o JAudio[5]. Outra ferramenta disponibilizada recentemente é o ACE (*Autonomous Classifier Engine*)[6] (MCKAY et al., 2005) que tem como

[4]Base de dados com músicas obtidas de http://magnatune.com
[5]Disponível em: http://coltrane.music.mcgill.ca/ACE/features.html
[6]Disponível em: http://coltrane.music.mcgill.ca/ACE/

objetivo ser uma plataforma específica para realizar experimentos que permitem explorar o uso de diferentes métodos e técnicas de combinação de classificadores para tarefas relacionadas a MIR.

Com o intuito de criar uma base de dados pública para a tarefa, Homburg et al. (2005) disponibilizaram uma base de 1.886 músicas obtidas a partir do site Garageband. A única limitação desta base é que cada música é representada por uma amostra de 10 segundos extraído aleatoriamente da música. A base está dividida em 9 gêneros sendo: *Blues* (120); *Electronic* (113); *Jazz* (319); *Pop* (116); *Rap/HipHop* (300); *Rock* (504); *Folk/Country* (222); *Alternative* (145); *Funk/Soul* (47).

Desta forma, considerando a limitação das poucas bases publicamente disponíveis, no trabalho de (MCKAY; MCENNIS; FUJINAGA, 2006) é apresentada a CODAICH database que possui 20.894 músicas no formato MP3 de 1.941 artistas. Os detalhes da base podem ser acessados nos formatos: iTunes XML, ACE XML, Weka ARFF ou jMusicMetadata HTML files. As músicas são classificadas de acordo com 53 gêneros possíveis.

Porém, um dos principais problemas existentes após o desenvolvimento de bases de dados musicais é como distribuí-las para os demais pesquisadores por causa das questões de direitos autorais. No intuito de centralizar o acesso a diversas bases de dados, sem ferir as questões de direitos autorais, no trabalho de (MCENNIS; MCKAY; FUJINAGA, 2006) foi desenvolvido o OMEN (*On demand Metadata ExtractioN*) que é uma plataforma para centralizar o acesso às bases de dados que sejam criadas e para permitir o acesso a CODAICH database. Algumas questões levantadas neste trabalho é que permitir que todas as possibilidades de extração de características fossem previamente calculadas e disponibilizadas iriam gerar uma explosão combinatorial em termos de processamento e também em termos de recursos de armazenamento. Desta forma, é apresentada uma interface para o pesquisador que seleciona quais as características que deseja trabalhar e a forma como elas devem ser extraídas e isso é feito sob demanda para contornar as limitações anteriores. Em alguns casos é possível fazer o armazenamento temporário das características calculadas, quando há espaço para tal. Como para a extração de características é utilizado o JAudio, é possível através da interface submeter os códigos fontes em java para que outras características possam ser disponíveis na plataforma.

2.4.5 Críticas à Tarefa

Com a atenção recebida pela tarefa de classificação automática de gêneros musicais, Aucouturier e Pachet (2003) fizeram um *survey* sobre essa tarefa. Neste trabalho eles descrevem experimentos no sentido de definir taxonomias para a tarefa. Porém, chegam à

conclusão que gêneros musicais são normalmente mal definidos (*ill-defined*), logo, sistemas que classificam baseados em gêneros são mal definidos, pois apresentam esta limitação. Eles classificam as abordagens para a classificação automática de gêneros em duas (por sinal, as mesmas que em qualquer sistema de RP): as treináveis e as baseadas em agrupamento. Nesse trabalho eles fazem um crítica aos sistemas baseados em janelas de análise por não utilizarem as informações temporais da música. Outro aspecto criticado é o baixo número de gêneros utilizados assim como a falta de métodos de seleção de características para gêneros específicos, pois para um determinado gênero musical as informações obtidas do timbre global da música podem não ser interessantes. Outro aspecto abordado é que não há padronização dos resultados nos trabalhos anteriores. Os autores ainda sugerem o uso de duas técnicas oriundas da área de mineração de dados conhecidas como Filtragem Colaborativa e Análise de Co-ocorrência para determinar a similaridade de músicas. Para a construção de novas bases de dados para o problema eles sugerem criar bases de dados utilizando compilações de músicas de um determinado ritmo (i.e. *Best of Italian Love Songs*).

No trabalho de (MCKAY; FUJINAGA, 2006) é feita uma análise crítica se a tarefa de classificação automática de gêneros musicais mereceria ou não continuar a ser pesquisada/tratada. Antes de apresentar os argumentos, eles utilizam a definição de (FABBRI, 1999) para definir os gêneros musicais como sendo: "um tipo de música, como ela é aceita por uma comunidade por qualquer razão, propósito ou critério". As principais conclusões apresentadas neste trabalho são:

1. Para aumentar o desempenho dos sistemas de classificação automática de gêneros musicais é necessário utilizar outros mecanismos além do timbre, como informações culturais disponíveis na *web*;

2. Possibilitar a atribuição de mais de um gênero para cada música, seja na saída do classificador, seja na rotulação da base de dados;

3. A aquisição de dados para *Ground truth* e sua respectiva classificação tem que ser considerados objetivos prioritários por si só;

4. Permitir uma estrutura, mesmo que simples, de ontologia mapeando as relações entre os gêneros;

5. Outra questão levantada considera que diferentes partes de uma música podem pertencer a diferentes gêneros, assim como podem ser representações diferentes do mesmo gênero e argumentam que utilizar as médias das características ao longo

de longas janelas de análise ou mesmo da música inteira pode ser uma abordagem limitadora;

6. De uma perspectiva musicológica, eles desencorajam o uso de técnicas como PCA para a redução de características, por mais que isto possa promover uma melhora na taxa de acerto. Isto limita a qualidade dos resultados de uma perspectiva teórica, pois são perdidas informações importantes como quais características são mais úteis em diferentes contextos, e sugerem o uso de mecanismos de seleção de atributos baseados em FFS, BFS e algoritmos genéticos;

7. Por fim, eles apontam para a necessidade de realizarem mais pesquisas no aspecto psicológico da classificação de gêneros musicais realizadas pelas pessoas considerando especialistas, não especialistas, pessoas de diferentes idades, culturas e experiências. Pois isto seria benéfico não apenas para melhorar o *ground truth* da área como também desenvolver diferentes sistemas para diferentes audiências e suas respectivas necessidades.

2.5 Avaliação Crítica

Um aspecto comum a maioria dos trabalhos da literatura é que eles estão normalmente propondo novos métodos de extração de características em conjunto com classificadores bem definidos. Como pode ser visto na proposta do ACE, mecanismos de combinação de classificadores foram pouco estudados e utilizados para a tarefa de reconhecimento automático de gêneros musicais. Outro aspecto que só recentemente tem sido investigado neste domínio é o uso de mecanismos de seleção de atributos.

Um outro aspecto importante é que as únicas bases disponíveis publicamente são a do trabalho de Tzanetakis e Cook (2002) (GTZAN), a base desenvolvida no trabalho de Homburg et al. (2005) e a CODAICH database (MCKAY; MCENNIS; FUJINAGA, 2006). Porém, as duas primeiras bases possuem sérias limitações: na primeira estão disponíveis apenas os primeiros 30 segundos de cada música no formato de áudio PCM. Na segunda estão disponíveis apenas 10 segundos extraídos de segmentos aleatórios de cada música. Com exceção dessas duas bases, as demais utilizadas na literatura possuem poucas músicas, e os gêneros utilizados são normalmente os mesmos (*Rock*, *Classical*) e os gêneros são disjuntos, ou seja, não existem trabalhos com subgêneros realmente próximos como *House* e *Trance*. No caso da terceira base, ela foi publicada somente em novembro de 2006, impossibilitando seu uso neste trabalho.

Dessa forma, tendo em mente o trabalho de Aucouturier e Pachet (2003), onde

é mostrado que definir uma taxonomia para gêneros é uma tarefa mal formulada, uma possível solução para este problema seria utilizar uma classificação um pouco mais abrangente baseada na percepção humana de como os gêneros são dançados. Apesar de não ser abrangente o suficiente para incluir todos os gêneros músicais possíveis, esta abordagem permitiria a construção de uma base de dados usando características culturais de diversos tipos de música.

Capítulo 3

Uma Proposta de Método para Classificação Automática de Gêneros Musicais

Como mostrado no capítulo anterior, a grande maioria dos trabalhos da área considera apenas o uso de um único segmento da música para realizar a classificação dos gêneros musicais. Além disto, as bases de dados existentes para realizar a tarefa possuem uma série de problemas e/ou limitações. Desta forma, para poder verificar as hipóteses deste trabalho, existe a necessidade do desenvolvimento de uma nova base de dados para a tarefa. O procedimento utilizado para a construção desta base é apresentado na seção 3.1. Na seção 3.2 é apresentada a abordagem para classificação automática de gêneros musicais, que consiste na extração de características de diferentes partes da música, o treinamento de um classificador para cada segmento e a combinação destes segmentos utilizando as estratégias de OAA, RR e baseadas nos escores de confiança produzidos por cada classificador. No intuito de melhorar os resultados de classificação individuais dos segmentos, e desta forma possivelmente melhorar a taxa de acerto dos mesmos, foram utilizados mecanismos de seleção de atributos utilizando AG's, que são apresentados na seção 3.3.

3.1 Criação e Manutenção da Base de Dados

Tendo em vista as limitações das bases desenvolvidas nos trabalhos anteriores para a verificação das hipóteses deste trabalho, surgiu a necessidade do desenvolvimento de uma nova base de dados para a tarefa. Porém considerando o esforço humano necessário para fazer a atribuição manual de gêneros as músicas, e também que uma base desenvolvida com cuidado poderia ser utilizada em outras tarefas além da classificação automática de gêneros musicais, foi necessário planejar como seria realizada a atribuição dos gêneros e o armazenamento, acesso e recuperação dessas informações.

Antes de iniciar o processo de aquisição, classificação e armazenamento das músicas, foi definido que seriam adquiridas pelo menos 3.000 músicas de 10 gêneros distintos de forma a poder fazer uma contribuição real para a área, visto que até então a base de dados mais abrangente (GTZAN) era composta por 1.000 músicas (limitadas a apenas os primeiros trinta segundos) de 10 gêneros.

3.1.1 O Processo de Atribuição de Gêneros Musicais

Neste trabalho o processo utilizado para atribuir um gênero a cada música é baseado na percepção humana de como cada música é utilizada para a dança. Para realizar este processo foram consultados dois profissionais com mais de dez anos de experiência no ensino de danças de salão. Estes profissionais fizeram uma primeira seleção das músicas que eles julgavam pertinentes a um determinado gênero de acordo com a forma que este era dançado e o autor deste trabalho verificou cada uma das músicas inicialmente selecionadas para evitar que equívocos fossem cometidos devido ao desgaste produzido pelo esforço humano necessário para realizar a tarefa. Em média foram classificadas 300 músicas por mês, sendo que o processo total para a criação da base de dados demorou um ano.

Como resultado desse esforço, foi desenvolvida a *Latin Music Database* que conta com 3.160 músicas de 10 gêneros musicais. Os gêneros musicais disponíveis na base e respectivos números de músicas são: *Tango* (404); *Salsa* (303); *Forró* (315); *Axé* (304); *Bachata* (308); *Bolero* (302); *Merengue* (307); *Gaúcha* (306); *Sertaneja* (310); *Pagode* (301). No total a base possui 543 artistas diferentes.

É importante ressaltar que na base desenvolvida foi utilizado este protocolo de inspeção humana de acordo com como as músicas são utilizadas para a dança. Ao contrário do que foi sugerido no trabalho de Aucouturier e Pachet (2003) para utilizar CDs de coleções completas, no caso dos rítmos latinos esta abordagem se mostrou ineficiente. Por exemplo, no caso da coletânea de quatro CDs (*Los 100 Mayores Exitos De La Musica Salsa*) apenas metade (50 das 100) das músicas podem ser classificadas como *Salsa*, as demais músicas desta coletânea são de outros gêneros musicais como Merengue, Lambada, Zouk e até mesmo Samba. Outra opção teria sido basear a classificação de todas as trilhas de um determinado álbum de acordo com o perfil do artista. Desta forma todas as músicas de Carlos Gardel seriam classificados como Tango. Porém, é importante ressaltar, que de todas as suas mais de 500 composições apenas cerca de 400 são *Tangos*. Desta forma introduziria ruído desnecessário na base. Por este motivo todas as músicas utilizadas nesta base foram avaliadas manualmente uma a uma e somente aquelas que realmente pertencem aos gêneros em questão foram rotulados como sendo desses gêneros. E mesmo

no caso de outros artistas de um determinado gênero, como *Salsa*, muito dificilmente todas as trilhas de seus álbuns são apenas *Salsas*.

Ao longo do processo de criação da base foi observado que normalmente cerca de uma a três músicas não são do gênero principal do perfil do artista.

3.1.2 Armazenamento, Acesso e Recuperação das Músicas

Além da aquisição das músicas e suas repectivas atribuições de gênero, para o desenvolvimento da base e sua ampla utilização em outras tarefas, várias reflexões foram realizadas no sentido de: criar uma base que possa ser facilmente utilizada para outras tarefas; permitir total reprodutibilidade dos experimentos realizados; evitar duplicidade das músicas cadastradas; facilitar o registro de novas músicas e/ou novos gêneros. Desta forma, tendo em mente estas várias questões, nesta seção são apresentadas as soluções adotadas para atingir esses objetivos.

O processo de armazenamento de uma nova música na base ocorre da seguinte forma:

1. Atribuição de um gênero à música em questão seguindo o procedimento descrito na SubSeção 3.1.1;

2. Inspeção manual do rótulo ID3 da música para verificar se os campos estão preenchidos corretamente e também de corrigí-los/adaptá-los a um padrão simples que consiste na padronização dos nomes e no uso do caracter especial & para indicar o nome de mais de um artista na mesma música. Os campos obrigatórios para cadastrar uma nova música são o *Artista* e o *Título* da música. A razão para essa abordagem é simples, mesmo que apenas uma pessoa esteja trabalhando no cadastro de músicas na base de dados, eventualmente álbuns do mesmo artista conterão trilhas com músicas presentes em outros álbuns, como por exemplo, no caso de um álbum com os maiores sucessos de um artista. Desta forma, este procedimento permite evitar duplicidade de músicas interpretadas pelo mesmo artista na base. Este controle de duplicidade é realizado no sistema quando uma nova música é cadastrada.

3. Cadastramento da música no sistema. Nesta etapa o sistema obtém os dados da música, verifica se não há duplicidade, atribui um código identificador para a música, associa esta música ao gênero pré-determinado e cria uma cópia da música. A informação do gênero da música é armazenada no banco de dados, pois como visto anteriormente, o campo *Genre* dos rótulos ID3 não é confiável. Além disto, no caso

de trabalhos futuros onde seja necessário o uso de alguma hierarquia, esta modificação pode ser incorporada facilmente ao sistema. No momento do cadastramento o sistema gera uma cópia da música cadastrada em um diretório pré-determinado seguindo a seguinte convenção:

```
DIRETORIO_GENERO\ARTISTA - TITULO - ALBUM - TRACK.MP3
```

onde DIRETORIO_GENERO é um diretório com o nome do gênero associado à música, e ARTISTA, TITULO, ALBUM e TRACK são informações obtidas do rótulo ID3 da música no momento em que ela é cadastrada.

O acesso à base de dados pode ser feito de forma convencional através do sistema de arquivos do sistema operacional, pois como mostrado, o sistema utiliza uma estrutura de arquivos e algumas regras de convenção simples para cadastrar as músicas. Porém, visando facilitar o acesso dos descritores das músicas pelos algoritmos de aprendizagem de máquina e também a reprodutibilidade dos experimentos, foram desenvolvidos dois módulos no sistema. Um para a extração de características e seu respectivo armazenamento no sistema e outro para a obtenção destas características já no formato utilizado por ferramentas de aprendizagem de máquina como é o caso do formato arff utilizado pelo WEKA (WITTEN; FRANK, 2005).

No que diz respeito à reprodutibilidade dos experimentos, com esta abordagem, todas as músicas disponíveis na base de dados têm as informações de Artista e Título. Com estas informações é possível criar junto com os arquivos arffs, gerados para os experimentos, uma lista das músicas utilizadas na mesma ordem em que elas serão utilizadas pelo módulo de classificação. O arquivo utilizado para armazenar esta lista é chamado de SAL (*Song Artist List*). O SAL é uma forma melhor de representar esta informação por três motivos:

1. Algumas vêzes artistas diferentes interpretam as mesmas músicas (porém, às vêzes, até mesmo em ritmos diferentes). Logo, utilizar apenas o Título da música não é suficiente;

2. Utilizar o ID da música fornecido pelo sistema não é confiável, pois se por algum motivo for necessário recadastrar todas as músicas, elas dificilmente serão cadastradas na mesma ordem em que foram cadastradas originalmente;

3. Pode ser que ao observar a lista das músicas utilizadas seja mais fácil de interpretar os resultados obtidos.

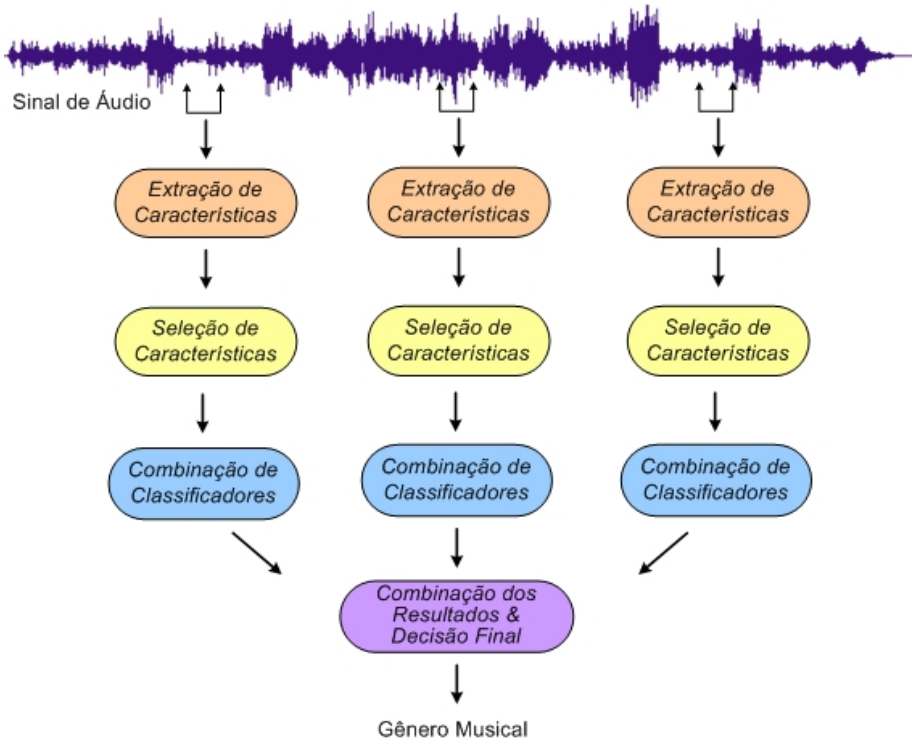

Figura 3.1: Visão Geral do Método Proposto

Um módulo para extração das características e seu armazenamento em banco de dados é uma opção interessante não apenas visando a reprodutibilidade dos experimentos, mas também em relação ao tempo que demora para calcular as características de cada música. Além disto, se em experimentos forem utilizados conjuntos de características diferentes das usadas neste trabalho, esta modificação exigiria apenas a adição de novas colunas nas tabelas existentes, permitindo uma comparação direta entre os resultados deste trabalho com as demais estratégias sendo propostas.

3.2 Método Para o Reconhecimento Automático

Uma visão geral do método proposto é apresentado na figura 3.1. O método proposto consiste na evolução dos trabalhos de Costa, Valle Jr. e Koerich (2004) e Koerich e Poitevin (2005). Além do método original proposto baseado na segmentação da música em três trechos (início, meio e fim), o acerto preditivo dos classificadores pretende ser melhorado com o uso dos métodos de decomposição do espaço do problema e de algoritmos de seleção de atributos.

3.2.1 Segmentação do Sinal de Áudio (Decomposição Temporal)

3.2.1.1 Definição

Convencionalmente, o problema da classificação automática de gêneros musicais pode ser definido como: dado um sinal de áudio de uma música S representado por um vetor de características D-dimensional, deseja-se atribuir uma classe (no caso um gênero musical) $g \in \mathcal{G}$ que melhor representa o vetor de características extraído de S. $\mathcal{G}$ é o conjunto de todos os gêneros musicais possíveis.

Em um problema típico de reconhecimento de padrões, dado um padrão de entrada, um vetor de características D-dimensional X_1^D é extraído de todo o padrão. Contudo,o sinal da música pode ser visto como um padrão variante no tempo. Desta forma, uma das soluções possíveis para levar em conta esta variabilidade intrínseca é extrair características do sinal da música inteira. Desta forma as características vão ser computadas ao longo do sinal. Contudo, extrair características da música inteira é um processo computacionalmente caro que deve ser evitado. Também não existe nenhuma garantia de que as características extraídas serão mais confiáveis do que outras abordagens que consideram características extraídas apenas de uma parte do sinal da música.

Por este motivo, a maior parte das abordagens para a classificação automática de gêneros musicais faz a extração de características de um número limitado de janelas da música. A maior desvantagem deste tipo de abordagem é que os valores das características se tornam dependentes dos quadros da música. Desta forma, estes valores variam de acordo com a posição das janelas utilizadas. Isto acontece porque a maioria das características proposta para a tarefa são variantes no tempo. A Figura 3.2 ilustra a variabilidade dos valores das características em relação à posição dos frames de onde elas foram extraídas. Na Figura 3.2 início, meio e fim representam os vetores de características extraídos destes trechos das músicas.

Neste trabalho é utilizada a estratégia de segmentação proposta no trabalho de Costa, Valle Jr. e Koerich (2004) ao invés de utilizar o início da música (TZANETAKIS; COOK, 2002) ou a música inteira (GRIMALDI; CUNNINGHAM; KOKARAM, 2003b). A estratégia de segmentação consiste em extrair segmentos do sinal de áudio e tomar a decisão baseada na combinação dos classificadores treinados e especializados para classificar cada um dos segmentos individualmente.

Formalmente esta abordagem pode ser definida como: sendo um sinal de áudio digital definido como uma seqüência $S =< s(1), s(2), \ldots, s(A) >= s_1^N$ onde $s(i)$ representa o sinal amostrado no instante i, e A é o número total de amostras que forma o sinal de áudio digital.

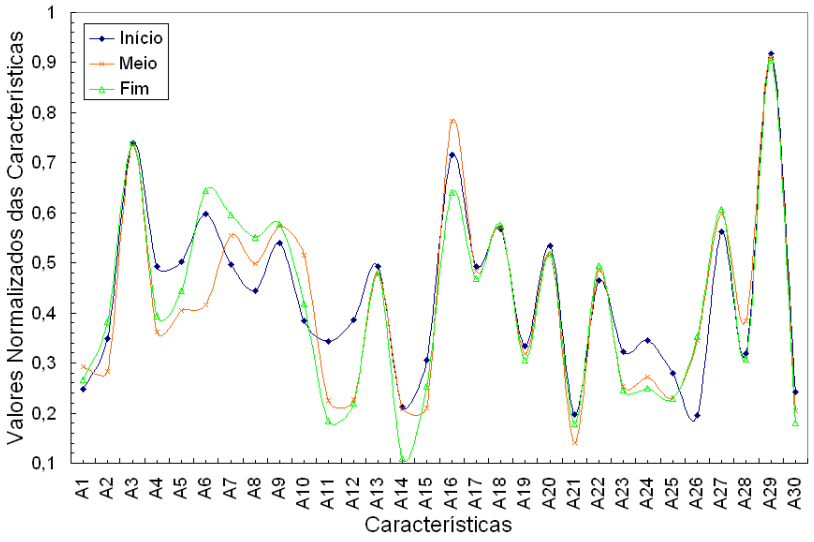

Figura 3.2: Média dos valores de 30 características extraídos de diferentes segmentos utilizando 150 músicas do gênero musical latino conhecido como Salsa.

Considerando que o sinal de áudio S é amostrado de acordo com uma freqüência f de amostras por segundo, e que o procedimento de extração de características é efetuado de acordo com uma janela de duração de t_w segundos, e que esta operação é realizada num intervalo de t_m segundos. Isto implica que existem ft_w amostras de áudio em cada segmento.

Desta forma, cada segmento de um sinal digital j é composto pelas amostras

$$f(j) = s(j.t_m.f + \hat{k}), \text{ onde } \hat{k} = 0, 1, \ldots, t_w.f - 1 \tag{3.1}$$

Ou seja, o primeiro segmento $f(0)$ considera amostras da música $< s(0), s(1), \ldots, s(t_w.f - 1) >$, e o quadro j^{th} engloba as amostras da música $f(j)$ em $< s(j.t_m.f), s(j.t_m.f + 1), \ldots, s(j.t_m.f + t_w.f - 1) >$. Destarte, para considerar a variação temporal ao longo do sinal de áudio, os vetores de características são obtidos realizando o procedimento de extração de características para cada segmento $f(j)$.

3.2.1.2 Aplicação

Nos experimentos realizados neste trabalho, do sinal de áudio da música S são extraídos três segmentos de trinta ($t_w = 30$) segundos. A principal razão para o uso de três segmentos ao invés de dois, quatro, cinco ou qualquer outro número é que este ainda é um problema em aberto pois além da segmentação do sinal da música em q quadros ainda existem outros problemas relacionados a segmentação como o tamanho da janela

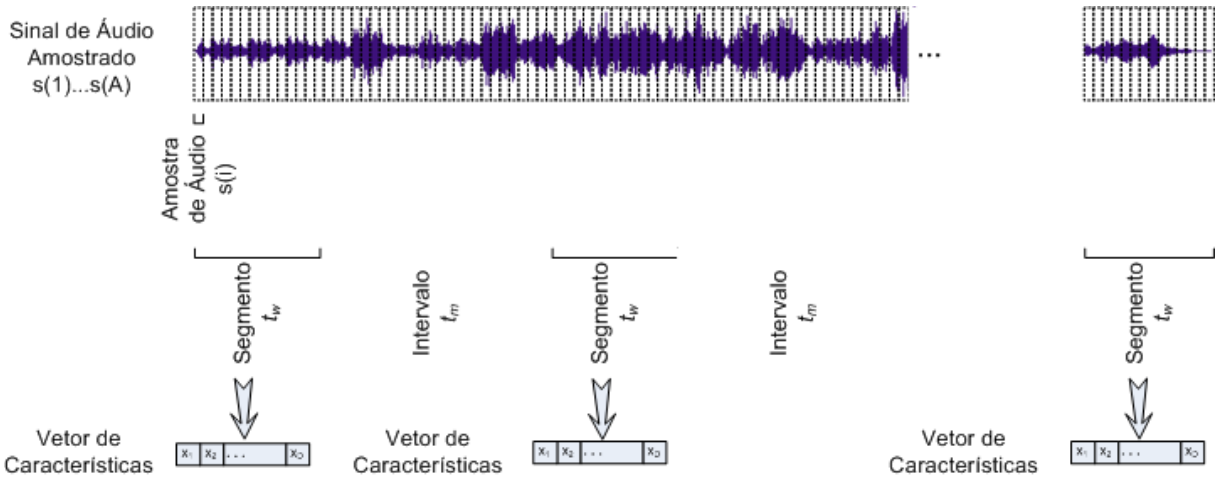

Figura 3.3: Visão geral do processo de extração de características

de análise utilizada. Este último problema foi investigado recentemente no trabalho de West e Cox (2005) onde eles avaliaram o desempenho de diferentes janelas de análise e propuseram uma técnica de segmentação automática. Outra razão para o uso de três segmentos é que, ao construir a base de dados, devido à natureza dos gêneros utilizados, normalmente foi necessário ouvir o meio da música e às vezes o final também para atribuir o gênero corretamente.

Lembrando que cada segmento da música é representado por $f(j)$, por uma questão de simplicidade, ao longo deste trabalho esta notação vai ser substituída por Seg_{parte}. Outra questão importante é que ao invés de utilizar um intervalo constante com um valor de (t_m) pré-definido para cada segmento, como estão sendo utilizados três segmentos, foi utilizada uma estratégia alternativa para definir o início de cada segmento. O ponto inicial dos segmentos é representando por t_{w_i}.

- Seg_{beg} representa o início da música. Neste segmento são utilizados os primeiros trinta segundos do sinal de áudio da música.

- Seg_{mid} representa o meio da música. Neste segmento são utilizados os trinta segundos do meio da música. Como a duração das músicas é variável, a estratégia utilizada para determinar o valor de t_{w_i} é a seguinte: o ponto inicial vai ser definido por: $t_{w_i} = (\frac{d}{3}) - 13$ segundos. Lembrando que d é a duração total da música.

- Seg_{end} representa o final de uma música. Entretanto para evitar pegar o final ruidoso ou silencioso que existe em algumas músicas no formato MP3, a estratégia utilizada para determinar o valor de t_{w_i} é: $i = d - 38$ segundos.

3.2.2 Extração de Características

Neste trabalho é utilizado o *framework* Marsyas para extrair características de diferentes segmentos do sinal de áudio e gerar vetores de características. O Marsyas implementa o conjunto de características propostas originalmente por Tzanetakis e Cook (2002) e utilizado em outros trabalhos (KOSINA, 2002) (LI; OGIHARA; LI, 2003). São considerados três tipos de características: relacionadas ao espectro sonoro (*Timbral texture*), relacionadas ao padrão rítmico (*beat-related*) e relacionadas à altura da nota (*pitch-related*). Características do espectro sonoro incluem a média e a variância do centróide espectral, do *rolloff* espectral, do fluxo espectral, das taxas de cruzamento zero, MFCC (*Coeficientes Cepstrais de Freqüência-Mel*), e da baixa energia. Características relacionadas ao padrão rítmico incluem as amplitudes relativas e as batidas por minuto. As características relacionadas ao *pitch* incluem os períodos máximos do pico do *pitch* nos histogramas. Estas características formam vetores de trinta dimensões (Espectro sonoro: 9 STFT + 10 MFCC; Padrão Rítmico: 6; *Pitch: 5*) que posteriormente são utilizados no treinamento de classificadores de maneira supervisionada. A seguir são descritas as características extraídas das músicas:

3.2.2.1 Características Relacionadas ao Espectro Sonoro

Centróide Espectral (*Spectral Centroid*) é o ponto balanceado do espectro. É uma medida da forma espectral e é associado freqüentemente com a noção do brilho espectral. O centróide espectral pode ser calculado como apresentado na equação 3.2.

$$C_t = \frac{\sum_{n=1}^{N} M_t[n] * n}{\sum_{n=1}^{N} M_t[n]} \tag{3.2}$$

onde $M_t[n]$ é o valor da transformada de Fourier no quadro t e faixa de freqüência n. O centróide espectral é um atributo perceptual importante na caracterização do timbre musical de instrumentos.

Rolloff Espectral (*Spectral Rolloff*) é outra medida da forma espectral que é definida como a freqüência R_t apresentada na equação 3.3 na qual 85% da magnitude da distribuição está concentrada.

$$\sum_{n=1}^{R_t} M_t[n] = 0.85 \sum_{n=1}^{N} M_t[n] \tag{3.3}$$

Fluxo Espectral (*Spectral Flux*) é uma medida da mudança espectral local e é definido como apresentado na equação 3.4.

$$F_t = \sum_{n=1}^{N} (N_t[n] - N_{t-1}[n])^2 \tag{3.4}$$

onde $N_t[n]$ é o valor normalizado da transformada de Fourier na janela t.

Taxas de Cruzamento Zero (*Time Domain Zero–Crossings*) é uma característica que ocorre quando as amostras sucessivas têm sinais diferentes. É calculada como apresentada na equação 3.5.

$$Z_t = \frac{1}{2} \sum_{n=1}^{N} |sign(x[n]) - sign(x[n-1])| \tag{3.5}$$

onde $x[n]$ é o sinal no domínio do tempo e a função *sign* é 1 ou 0 para os argumentos positivos e negativos respectivamente. Ao contrário do centróide espectral, do *rolloff* espectral e do fluxo espectral, que são características no domínio da freqüência, a taxa do cruzamento zero é uma característica no domínio do tempo.

Coeficientes Cepstrais da freqüência Mel (*Mel-frequency cepstral coefficients*) são características perceptualmente motivadas que também são baseadas na STFT (*Short Time Fourrier Transform*). Após obter a amplitude logarítmica da magnitude do espectro, as faixas pré-determinadas são agrupadas e suavizadas (*smoothed*) de acordo com a motivação perceptual da escala da freqüência Mel. Finalmente, para descorrelacionar os vetores de características resultantes, uma transformada discreta de coseno é utilizada. Apesar de normalmente treze coeficientes serem utilizados para representar a fala, experimentos mostram que os cinco primeiros coeficientes levam a um melhor desempenho para a classificação de gêneros musicais (TZANETAKIS; COOK, 2002).

Análise e Janela de Textura. Em análise de áudio o sinal é quebrado em pequenos segmentos de tempo sobrepostos e cada segmento é processado separadamente. Estes segmentos são chamados de janela de análise e devem ser pequenos o suficiente para que as características de freqüência do espectro de magnitude sejam relativamente estáveis. Entretanto a sensação de textura do som surge como resultado de múltiplos espectros de tempo curto com diferentes características seguindo algum padrão no tempo. Por exemplo, a fala contém vogais e consoantes as quais tem diferentes características espectrais.

Logo, de forma a capturar a longa natureza da textura do som, as características computadas são médias e variâncias das características descritas anteriormente nesta seção, em um número de janelas de análise. O termo janela de textura é utilizado para descrever esta janela maior e idealmente deve corresponder ao mínimo de tempo de som que é necessário para identificar a textura de um som ou de uma música. Essencialmente, ao invés de usar os valores das características diretamente, são calculados os parâmetros

de uma distribuição gaussiana multidimensional. Mais especificamente, os parâmetros (médias, variâncias) são calculados com base na janela de textura que consiste no vetor de características atual em adição a um número específico de vetores de características do passado.

Baixa Energia (*Low Energy*) é calculada sobre um número de janelas com a média e variação, e não separadas para cada janela como as outras características. A característica energia baixa é definida como a porcentagem das janelas que têm menos energia do que a energia média de todas as 40 janelas. Por exemplo, sinais musicais terão energia mais baixa que sinais de fala que normalmente contêm muitas janelas silenciosas.

Com as características apresentadas nesta seção, o espectro sonoro de uma música consiste nas seguintes características: médias e variâncias do centróide espectral, do *rolloff* espectral, do fluxo espectral, das taxas de cruzamento zero sobre a janela da textura (8), baixa energia (1) e as médias e variâncias dos cinco primeiros coeficientes MFCC sobre a janela de textura resultado assim em um vetor de características com dezenove dimensões.

3.2.2.2 Características Relacionadas ao Padrão Rítmico (Beat-Related)

A batida e a estrutura rítmica de uma música é freqüentemente uma boa indicação do gênero. Por exemplo, *dance music* tende a ter uma batida principal muito forte e distintiva. A música clássica, geralmente não tem uma batida dominante e regular desobstruída, devido à complexidade do arranjo. A extração de características da batida tenta encontrar a batida principal da música e de seu período em BPM (*Batidas Por Minuto*). Além desta, é calculada também à batida mais forte, e um número de características relacionando a primeira e segunda batida.

Inicialmente o sinal é decomposto em um número de bandas de freqüências usando uma transformada Wavelet discreta (KOSINA, 2002 apud SWELDENS; PIESSENS, 1993). Após esta decomposição, uma série de passos para a extração do envelope da amplitude no domínio do tempo é aplicada a cada banda: retificação de onda completa, filtragem passa-baixa, sub-amostragem e remoção das médias (KOSINA, 2002; TZANETAKIS; COOK, 2002).

Após o passo da extração, os envelopes de cada banda são somados e a autocorrelação resultante é calculada. Este resultado é uma função de autocorrelação onde os picos (*peaks*) dominantes correspondem ao tempo de *lag* (*time lags*) onde o sinal tem a auto-similaridade mais forte. Os primeiros três picos da função de autocorrelação são adicionados ao histograma de batida. Cada banda do histograma corresponde a um período da batida em BPM. Para cada um dos três picos selecionados, a amplitude do pico é

adicionada ao histograma. Este procedimento é repetido para cada janela de análise. Os picos mais fortes no final do histograma correspondem às batidas mais fortes do sinal. Seis características são calculadas usando o histograma de batidas:

- A amplitude relativa (i.e. a amplitude dividida pela soma de amplitudes) do primeiro e do segundo picos no histograma de batidas. Esta é uma medida de quão distintivas são as batidas comparadas com o resto do sinal.
- A razão da amplitude do segundo pico dividida pela amplitude do primeiro pico. Essa característica expressa a relação entre a batida principal e a primeira batida auxiliar.
- O período do primeiro e segundos picos em BPM, indicando quão rápida é a música.
- A soma do histograma, a qual pode ser um indicador da força da batida. A soma das bandas do histograma é uma medida de força da auto-similaridade entre as batidas, a qual é um fator de quão rítmica uma música parece ser.

3.2.2.3 Características Relacionadas à Altura da Nota (Pitch-Related)

O conjunto de características de conteúdo *pitch* é baseado em múltiplas técnicas de detecção de *pitch*. Neste algoritmo, o sinal é decomposto em duas bandas de freqüência (abaixo e acima de 1.000 Hz) e envelopes de amplitude são extraídos para cada banda da freqüência. A extração do envelope é realizada aplicando retificação de meia onda e filtro passa-baixa. Os envelopes são somados e uma função "aumentada" de autocorrelação é computada para que o efeito de múltiplos inteiros no pico das freqüências para múltiplos *pitch*'s detectados sejam reduzidos.

Os picos proeminentes desta função de autocorrelação "aumentada" correspondem aos principais *pitches* para aquele curto segmento de som. Esse método é similar a detecção da estrutura de batidas para curtos períodos correspondendo a percepção de *pitch*. Os três picos dominantes são acumulados em histogramas de *pitch* sobre todo o sinal de áudio. Para computar o histograma de *pitch*, é utilizada uma janela de análise de 512 amostras com taxa de amostragem de 22 050 Hz (aproximadamente 23 ms).

3.2.2.4 Vetor de Características Resultante

O vetor de características resultante é apresentando na Figura 3.4, onde é descrita a associação entre a posição no vetor e a característica relacionada. Um procedimento

Número da característica	Descrição
1–6	Características relacionadas à batida
7	Média do Centróide Espectral
8	Média do Rolloff Espectral
9	Média do Fluxo Espectral
10	Média das Taxas de Cruzamento Zero
11	Desvio Padrão do Centróide Espectral
12	Desvio Padrão do Rolloff Espectral
13	Desvio Padrão do Fluxo Espectral
14	Desvio Padrão das Taxas de Cruzamento Zero
15	Baixa Energia
16	Média do 1º MFCC
17	Média do 2º MFCC
18	Média do 3º MFCC
19	Média do 4º MFCC
20	Média do 5º MFCC
21	Desvio Padrão do 1º MFCC
22	Desvio Padrão do 2º MFCC
23	Desvio Padrão do 3º MFCC
24	Desvio Padrão do 4º MFCC
25	Desvio Padrão do 5º MFCC
26–30	Características relacionadas ao Pitch

Figura 3.4: Descrição do vetor de características

final que deve ser aplicado ao vetor de características resultante é um procedimento para normalização dos atributos para que esses possam ser utilizados pelos algoritmos de aprendizagem de máquina.

Desta forma a seguinte regra de normalização é utilizada: considerando MAX_VALUE como sendo o valor máximo do atributo e MIN_VALUE o valor mínimo, o novo valor do atributo (para cada instância) é dado por: NovoValor = (ValorAntigo - MIN_VALUE) / (MAX_VALUE - MIN_VALUE).

3.3 Seleção de Atributos

Como mostrado na figura 3.1 o mecanismo de seleção de atributos é aplicado a cada segmento, dessa forma o mecanismo de seleção de atributos pode ser avaliado independentemente em cada um dos vetores de características, denotados por $X_1 = (x_1 x_2 \dots x_{30}), X_2 = (x_1 x_2 \dots x_{30}), \dots, X_n = (x_1 x_2 \dots x_{30})$. A razão desse procedimento ser utilizado em todos os segmentos e não apenas em um único segmento e replicado (nesse caso utilizando as mesmas características) para os demais é porque em virtude do sinal da música ser variante no tempo, é possível que as características que melhor

representam o segmento do início da música (Seg_{beg}) não sejam as mesmas que melhor representam o fim da música (Seg_{end}).

O mecanismo de seleção de atributos utilizado neste trabalho consiste no uso da abordagem envelope em conjunto com o uso de AGs. Desta forma o conjunto original de características é utilizado para criar um conjunto de 20 indivíduos que utilizam a notação apresentada na seção 2.3. Esses indivíduos são gerados aleatóriamente, utilizando a notação mostrada anteriormente. Ou seja, são criados vetores binários de 30 posições onde 0 indica a ausência e 1 indica a presença da característica naquele indivíduo.

Desta forma, para cada indivíduo o seguinte procedimento vai ser aplicado:

1. É treinado um classificador específico para o indivíduo, que utiliza apenas as características indicadas como presentes;

2. O classificador é utilizado num conjunto de validação para calcular a função de adequabilidade do indivíduo, que é determinada pela taxa de classificação correta do classificador;

3. O valor de adequabilidade é retornado.

Após este procedimento ter sido aplicado a toda a população de indivíduos (neste trabalho é utilizada uma população de 50 indivíduos), vai ser verificado se a solução convergiu ou se o número máximo de gerações foi atingido. Se isto acontecer, é retornado como solução o indivíduo de maior adequabilidade, caso contrário são aplicadas as operações genéticas de mutação e crossover para gerar uma nova população de indivíduos e repetir o processo.

3.4 Classificação, Combinação e Decisão

As estratégias de combinação de classificadores utilizadas neste trabalho estão localizadas em dois níveis na figura 3.1. No primeiro nível a "combinação de classificadores" representa o uso das estratégias de decomposição do espaço do problema, criando uma abordagem híbrida baseada na decomposição de tempo (segmentação) e espaço (técnicas de decomposição do espaço do problema). Neste trabalho são utilizadas as técnicas de Um-Contra-Todos (*One-Against-All*) e *Round Robin*. No segundo nível, a combinação dos resultados dos classificadores individuais ou aliados às técnicas de decomposição do espaço do problema treinado em cada segmento vão ser combinadas através do voto da maioria ou de uma das regras baseada na probabilidade a posteriori fornecida pelos classificadores.

Desta forma, para cada segmento um classificador vai ser treinado. Este procedimento vai resultar em três classificadores: $C_{beg}, C_{mid}, C_{end}$. É importante ressaltar que o classificador de base utilizado é sempre homogêneo (ou seja, o mesmo classificador).

Outro aspecto importante é que a regra padrão utilizada no caso de empates que podem ocorrer ao utilizar as regras de votação da maioria e Max. De forma a resolver os empates e fornecer uma única saída, entre as classes que empataram, o gênero vai ser atribuído com base no escore de confiança mais alto. No caso de haver outro empate, desta vez com os escores de confiança, a decisão vai ser baseada nos segmentos que empataram. Se um dos votos empatados foi fornecido pelo classificador C_{mid}, o gênero vai ser decidido por ele. Se não, a decisão vai ser atribuída ao classificador C_{beg}.

Capítulo 4

Avaliação do Método Proposto

Neste capítulo são apresentados os experimentos realizados ao longo do desenvolvimento do trabalho. Em todos os experimentos apresentados neste capítulo foi utilizada uma versão completa da base *Latin Music Database* apresentada na seção 3.1. De forma a realizar os experimentos com um número balanceado de exemplos por classe, foram selecionadas 300 músicas de cada gênero. Para a extração de características foi utilizado o Marsyas (TZANETAKIS; COOK, 1999). Os valores apresentados nas tabelas de resultados são referentes à média dos valores obtidos utilizando validação cruzada fator 10.

É importante ressaltar que para evitar qualquer tipo de tendência (*bias*) nos experimentos, todas as músicas disponíveis foram selecionadas aleatoriamente sem repetição para compor os conjuntos utilizados nos experimentos. A questão da análise estatística dos resultados obtidos foi amplamente discutida durante o desenvolvimento do trabalho. Porém, como são utilizados três segmentos da música com o mesmo conjunto de características em todos os casos, não fazia sentido em comparar os diferentes segmentos da música, pois o objeto que estava sendo comparado não era o mesmo, apesar das características serem as mesmas. Se fosse possível realizar a análise estatística, teria sido utilizado um dos métodos abordados em (DEMSAR, 2006).

Outra observação importante é que o módulo de classificação do sistema é integrado ao *framework* WEKA (WITTEN; FRANK, 2005) que permite obter a probabilidade de cada classe mesmo para classificadores como o k-NN, que normalmente não fornece probabilidades na saída, sendo que neste caso as distâncias são convertidas em probabilidades.

4.1 Decomposição Temporal Vs. Música Inteira

Os objetivos deste primeiro experimento foram:

Tabela 4.1: Taxa de classificação correta (%) utilizando segmentos isolados vs. música inteira sobre o conjunto de testes.

	Taxa de Classificação Correta (%)			
Classificador	Seg_{beg}	Seg_{mid}	Seg_{end}	MI
J48	39.60	44.44	38.80	44.20
3-NN	45.83	56.26	48.43	57.96
MLP	53.96	56.40	48.26	56.46
NB	44.43	47.76	39.13	48.00
SVM	**57.43**	**63.50**	**54.60**	**63.40**

- Verificar o desempenho dos classificadores nos segmentos do início (Seg_{beg}), meio (Seg_{mid}) e fim (Seg_{end}) da música;

- Combinar classificadores utilizando o método de Decomposição Temporal com as diferentes regras de combinação baseadas nas probabilidades a posteriori de cada classificador;

- Verificar o desempenho deste método em relação aos classificadores treinados utilizando características extraídas da música inteira e aos classificadores que utilizam somente um segmento da música.;

A Tabela 4.1 apresenta os resultados obtidos por cada classificador treinado utilizando cada um dos três segmentos e também utilizando um classificador treinado utilizando as características extraídas da MI (*Música Inteira*). Uma análise dos resultados apresentados na tabela 4.1 mostra que considerando apenas os três segmentos, em todos os casos os melhores resultados são obtidos pelo Seg_{mid} e não pelo Seg_{beg} que é comumente utilizado na literatura. Outro aspecto interessante é que as características extraídas dos diferentes segmentos da música levam a resultados significativamente diferentes considerando a taxa de classificação correta. Alguns motivos para isto é que devido aos gêneros utilizados, como Salsa em que algumas vezes começa como uma música lenta e depois de algum tempo elas "explodem"e diversos instrumentos começam a tocar. Outra possível razão para este comportamento é que o Seg_{mid} é normalmente mais estável que o resto da música.

Ao comparar o desempenho obtido utilizando os segmentos individuais em relação à música inteira, observou-se que utilizar a música inteira fornece resultados similares ou superiores ao Seg_{mid}. Entretanto, utilizar características da música inteira é computacionalmente mais caro do que utilizar apenas um único segmento. Por exemplo, para uma música de 4 minutos e 57 segundos, para extrair características da música inteira são gastos em média 56 segundos e 40 segundos para extrair características dos três segmentos.

A tabela 4.2 apresenta os resultados para cada classificador utilizando diferentes regras de combinação a-posteriori onde, MAJ indica Voto da Maioria (*Majority Voting)*, WS indica Soma Ponderada (*Weighted Sum*). Para WS I (*Weighted Sum* I) foram considerados os valores de $\alpha = 0.3$, $\beta = 0.6$ e $\gamma = 0.1$. Para WS II foram considerados os valores de $\alpha = 0.25$, $\beta = 0.5$ e $\gamma = 0.25$. WP indica Produto Ponderado (*Weighted Product* e os pesos utilizados foram os mesmos que para WS I e II respectivamente. Os pesos foram definidos desta forma para atribuir ao classificador do Seg_{mid} um peso um pouco maior que os demais. Contudo, estes pesos poderiam ser melhorados utilizando algoritmos genéticos para buscar uma melhor combinação dos pesos. Este procedimento possivelmente levaria a um desempenho um pouco melhor. No intuito de facilitar a comparação dos métodos de combinação em relação aos respectivos classificadores utilizando a MI, os resultados da tabela 4.1 são repetidos.

Tabela 4.2: Taxa de classificação correta (%) utilizando várias regras para a combinação de classificadores vs. música inteira sobre o conjunto de testes.

	Taxa de Classificação Correta (%)								
Classificador	MAJ	MAX	SUM	WS I	WS II	PROD	WP I	WP II	MI
J48	47.33	43.76	47.30	45.93	**47.63**	20.50	20.50	20.50	44.20
3-NN	60.46	60.67	**62.66**	61.60	63.16	62.06	61.40	62.26	57.96
MLP	59.43	57.40	61.83	58.16	59.63	62.50	62.16	**62.56**	56.46
NB	46.03	45.96	46.66	**48.40**	47.80	46.13	48.06	46.80	48.00
SVM	65.06	64.13	65.73	**66.76**	66.46	65.50	66.50	66.30	63.40

A análise dos resultados da tabela 4.2 mostram que a eficiência das regras de combinação dependem dos classificadores utilizados. Para os classificadores 3-NN, MLP e SVM, em todos os casos as regras de combinação fornecem melhores resultados do que utilizar a música inteira. Para o Classificador NB as únicas regras de combinação que provêm resultados similares ao utilizar a música inteira são a soma e produto ponderados I (WS I e WP I respectivamente). Para o J48 as regras de voto da maioria, soma e somas ponderadas apresentam resultados melhores do que a música inteira. O baixo desempenho obtido pelo classificador J48 em relação à regra de combinação do produto e produto ponderado é em função de um grande número de empates que ocorreram no caso deste classificador. Este foi o único classificador em que ocorreram empates utilizando as regras de combinação. Os melhores resultados em todos os experimentos desta seção foram obtidos utilizando o classificador SVM. A melhor taxa de classificação obtida foi de 66.76% utilizando a regra de combinação WS I.

Com os experimentos realizados nesta seção ficou provado que o uso da estratégia de decomposição temporal fornece melhores resultados do que utilizar a música inteira ou

os segmentos individualmente.

4.2 Decomposição Temporal–Espacial

Os objetivos deste segundo experimento foram:

- Verificar o desempenho dos classificadores nos segmentos do início (Seg_{beg}), meio (Seg_{mid}), fim (Seg_{end}) e na música inteira utilizando os métodos de Decomposição Espacial, mais especificamente OAA e RR;
- Utilizar a estratégia de Decomposição Temporal–Espacial e comparar seu desempenho com o método de Decomposição Temporal;

A Tabela 4.3 apresenta os resultados obtidos por cada classificador treinado utilizando cada um dos três segmentos com as estratégias de OAA e RR. De forma a facilitar a comparação destes métodos aos resultados obtidos nos experimentos da Seção 4.1, a coluna BL (*Baseline*) apresenta os resultados obtidos por cada classificador em cada segmento sem utilizar os métodos de decomposição do espaço do problema para cada segmento da música. Também é importante ressaltar que como o classificador SVM foi utilizado por padrão com a estratégia de decomposição *Round Robin*, os resultados para a coluna BL foram omitidos.

Tabela 4.3: Taxa de Reconhecimento (%) utilizando OAA e RR nos segmentos individuais.

	Seg_{beg}			Seg_{mid}			Seg_{end}		
Classificador	BL	OAA	RR	BL	OAA	RR	BL	OAA	RR
J48	39.60	41.56	45.96	44.44	44.56	49.93	38.80	38.42	45.53
3-NN	45.83	45.83	45.83	56.26	56.26	56.26	48.43	48.43	48.43
MLP	53.96	52.53	55.06	56.40	53.08	54.59	48.26	51.96	51.92
NB	44.43	42.76	44.43	47.76	45.83	47.79	39.13	37.26	39.19
SVM	–	26.63	57.43	–	36.82	63.50	–	28.89	54.60

A análise dos resultados apresentados na tabela 4.3 mostra que para o classificador J48 a estrátegia RR melhora os resultados do classificador enquanto que a estrátegia de OAA produz melhoras mas não tão significativas. Para o classificador 3-NN os resultados são sempre os mesmos, porém uma observação interessante olhando os arquivos de resultados é que as respostas fornecidas em cada estratégia são diferentes. Para a MLP as estratégias de decomposição melhoram o desempenho do início da música utilizando RR e no final da música utilizando tanto OAA como RR. Para o classificador NB as estratégias de decomposição não melhoraram a performance do classificador. Para o SVM pode ser

observado como a escolha da estratégia de decomposição do espaço do problema pode fazer com que o melhor classificador utilizando RR possa ser o pior ao utilizar a estratégia de OAA. Em geral, o método de RR fornece resultados superiores ou equivalentes ao BL, a única exceção foi para o classificador MLP para o Seg_{mid}.

Na tabela 4.4 são apresentados os resultados utilizando as técnicas de decomposição do espaço do problema na música inteira, e também com o método de Decomposição Temporal–Espacial. Nessa tabela os valores do BL são definidos pelos valores obtidos sem estratégias de decomposição do espaco do problema para música, e utilizando os valores do experimento anterior com o voto da maioria.

Tabela 4.4: Taxa de reconhecimento (%) utilizando decomposição temporal–espacial vs. música inteira

	Ensembles			MI		
Classificador	BL	OAA	RR	BL	OAA	RR
J48	47.33	49.63	54.06	44.20	43.79	50.63
3-NN	60.46	59.96	61.12	57.96	57.96	59.93
MLP	59.43	61.03	59.79	56.46	58.76	57.86
NB	46.03	43.43	47.19	48.00	45.96	48.16
SVM	–	30.79	65.06	–	37.46	63.40

A análise dos resultados apresentados na tabela 4.4 mostra que para a música inteira o uso da estrátegia de RR sempre aumenta a taxa de acerto de qualquer um dos classificadores em relação ao BL, enquanto a estratégia de OAA fornece resultados superiores apenas para a rede neural MLP. Ao comparar os resultados da MI em relação aos três segmentos, a música inteira utilizando RR sempre fornece resultados melhores do que os obtidos por qualquer segmento indiferente da estratégia de combinação utilizada. A única exceção é para a rede MLP em que os melhores resultados são obtidos utilizando a estratégia de OAA.

No caso dos resultados utilizando Decomposição Temporal–Espacial, as estratégias de OAA e RR podem melhorar o desempenho dos classificadores, mas esta melhora não é significativa. Já a comparação da música inteira em relação ao método de Decomposição Temporal–Espacial mostra um resultado semelhante ao experimento anterior, visto que para os classificadores J48, 3-NN e MLP em todos os casos as regras de combinação fornecem melhores resultados do que utilizar a música inteira e para o NB os resultados são inferiores ou similares. Para o SVM os resultados só são superiores ao utilizar a estratégia de RR.

4.3 Seleção de Características

Os objetivos deste terceiro experimento foram:

- Utilizar o método de seleção de atributos com algoritmos genéticos em cada um dos segmentos utilizados pelo método de Decomposição Temporal e também na música inteira;
- Verificar se a combinação das características selecionadas, pelo método de seleção de atributos, em cada segmento em conjunto com as diferentes regras de combinação, baseadas nas probabilidades a posteriori de cada classificador, aumentam a taxa de reconhecimento dos classificadores;
- Verificar se o método de seleção de atributos consegue discriminar um subconjunto de características relevantes para aumentar a taxa de reconhecimento dos classificadores treinados com características extraídas da música inteira.
- Comparar o desempenho dos classificadores utilizando seleção de atributos com Decomposição Temporal e diferentes regras de combinação baseadas nas probabilidades a posteriori em relação aos classificadores treinados utilizando seleção de atributos com a música inteira.

A Tabela 4.5 apresenta os resultados obtidos por cada classificador treinado utilizando cada um dos três segmentos utilizando o número de características selecionadas pelo AG (que são indicadas na coluna #) e também os resultados obtidos sem utilizar a seleção de atributos que são tidos como *Baseline* (BL).

Tabela 4.5: Taxa de classificação correta (%) utilizando seleção de atributos com AG nos segmentos individuais.

	Seg_{beg}			Seg_{mid}			Seg_{end}		
Classificador	BL	GA	#	BL	GA	#	BL	GA	#
J48	39.60	44.70	15	44.44	45.76	21	38.80	38.73	18
3-NN	45.83	51.19	14	56.26	60.02	18	48.43	51.11	19
MLP	53.96	52.73	22	56.40	54.73	24	48.26	47.86	18
NB	44.43	45.43	21	47.76	50.09	18	39.13	39.66	24
SVM	57.43	57.13	24	63.50	59.70	22	54.60	55.33	24

A análise dos resultados apresentados na tabela 4.5 mostra que o método de seleção de atributos se mostrou eficiente para os classificadores J48, 3-NN e NB melhorando a taxa de acerto ou pelo menos mantendo uma taxa de acerto similar, porém com um número

reduzido de atributos. Para a rede MLP o número de atributos foi reduzido porém a taxa de reconhecimento foi prejudicada. Isto também aconteceu com o SVM com exceção do Seg_{end}, onde o método de seleção de atributos forneceu um número menor de atributos com uma melhora na taxa de classificação.

Na tabela 4.6 são apresentados os resultados utilizando a técnica de Decomposição Temporal com os vetores de características gerados utilizando o método de seleção de atributosdos para cada segmento. Também são apresentados os resultados para a música inteira com o respectivo número de atributos selecionados (#). Os resultados obtidos utilizando o método de Decomposição Temporal sem o uso do método de seleção de atributos, foram repetidos para facilitar a comparação dos resultados.

Tabela 4.6: Taxa de classificação correta (%) utilizando várias regras para a combinação de classificadores vs. música inteira sobre o conjunto de testes.

Classificador	MAJ	MAX	SUM	WS I	WS II	PROD	WP I	WP II	MI
J48	47.33	43.76	47.30	45.93	47.63	20.50	20.50	20.50	44.20
J48-GA	50.10	46.00	50.90	47.03	49.30	22.83	22.83	22.83	45.13 (#19)
3-NN	60.46	60.67	62.66	61.60	63.16	62.06	61.40	62.26	57.96
3-NN-GA	63.20	63.00	63.93	64.43	64.86	64.26	64.80	64.73	57.30 (#16)
MLP	59.43	57.40	61.83	58.16	59.63	62.50	62.16	62.56	56.46
MLP-GA	59.30	57.70	60.33	56.90	58.33	60.90	60.73	60.96	53.56 (#24)
NB	46.03	45.96	46.66	48.40	47.80	46.13	48.06	46.80	48.00
NB-GA	47.10	44.80	48.00	51.46	51.03	47.76	51.06	49.36	49.63 (#19)
SVM	65.06	64.13	65.73	66.76	66.46	65.50	66.50	66.30	63.40
SVM-GA	63.03	60.43	64.40	64.40	64.40	64.90	64.60	64.36	61.13 (#22)

A análise dos resultados da tabela 4.6 mostra que para a música inteira, o método de seleção de atributos reduz o número de características porém apenas para os classificadores J48 e 3-NN a taxa de reconhecimento é similar ao resultados obtidos sem o uso do método de seleção de atributos. Já o método de Decomposição Temporal aliado à seleção de atributos fornece melhoras na taxa de reconhecimento dos classificadores J48 e 3-NN em todos os casos. Os classificadores MLP e NB possuem taxas de reconhecimento melhores em alguns casos e piores em outros e no caso do SVM o método utilizando seleção de atributos sempre piora ou é igual a taxa de reconhecimento. Em comparação com a música inteira, considerando apenas os resultados utilizando seleção de atributos, os resultados são similares aos do experimento 1, ou seja, para os classificadores 3-NN-GA e MLP-GA para todos os casos as regras de combinação fornecem melhores resultados do que utilizar a música inteira. Para o classificador SVM isto acontece em todos os casos, menos para a regra de combinação MAX. Para o J48, com exceção das regras baseadas no produto, todas as outras apresentam resultados melhores do que a música inteira,

Tabela 4.7: Características selecionadas para cada segmento dos classificadores 3-NN, J48 e MLP.

	3-NN				J48				MLP			
#	Full	Seg_{beg}	Seg_{mid}	Seg_{end}	Full	Seg_{beg}	Seg_{mid}	Seg_{end}	Full	Seg_{beg}	Seg_{mid}	Seg_{end}
1					X			X	X		X	
2									X		X	
3					X			X	X		X	
4						X	X	X			X	
5										X	X	
6	X	X	X	X	X	X		X	X	X	X	X
7	X		X	X	X	X	X		X	X	X	
8				X		X	X	X		X	X	
9	X	X	X	X	X	X	X	X	X	X	X	X
10		X	X		X		X	X	X	X	X	X
11	X	X			X			X	X	X	X	X
12			X	X			X		X	X	X	X
13	X	X	X	X	X		X	X	X	X	X	X
14				X	X		X		X	X		X
15	X		X	X	X	X	X	X	X	X	X	X
16	X	X	X	X	X	X	X	X	X	X	X	X
17	X	X	X	X		X			X	X	X	X
18	X	X	X	X	X	X	X	X	X	X	X	X
19	X	X	X	X			X		X	X	X	X
20	X		X	X	X		X		X	X	X	
21	X	X	X	X	X	X	X	X	X	X	X	X
22	X	X	X	X	X	X	X		X	X	X	X
23	X	X	X	X	X	X	X	X	X	X	X	X
24	X		X	X	X				X	X	X	X
25	X	X	X	X	X		X	X	X	X		X
26					X		X			X		
27						X	X	X	X			
28		X	X	X		X	X	X	X		X	X
29						X						
30							X	X				

enquanto que para o NB as regras com pesos ponderados (WS I, WS II, WP I, WP II) apresentam resultados superiores a utilizar a música inteira.

Desta forma o método de seleção de características considerando os segmentos individuais só se mostrou eficiente, melhorando a taxa de classificação e reduzindo o número de características, para os classificadores J48, 3-NN e NB. Utilizando a música inteira o método só foi eficente para os classificadores J48 e NB. A combinação do método

Tabela 4.8: Características selecionadas para cada segmento dos classificadores NB e SVM.

	NB				SVM			
#	Full	Seg_{beg}	Seg_{mid}	Seg_{end}	Full	Seg_{beg}	Seg_{mid}	Seg_{end}
1	X	X		X	X	X	X	X
2	X		X		X		X	X
3	X		X	X				
4		X	X	X				X
5		X		X	X	X		
6	X	X	X	X	X	X	X	
7						X		
8					X	X	X	X
9	X	X	X	X	X	X	X	X
10	X	X	X	X	X	X	X	X
11				X	X	X	X	
12	X		X	X	X	X	X	X
13	X	X	X	X	X	X	X	X
14							X	X
15	X	X	X	X	X	X	X	X
16	X	X	X	X	X	X	X	X
17	X	X	X	X	X	X	X	X
18	X	X	X	X	X	X	X	X
19	X	X	X	X		X	X	X
20		X		X	X	X		X
21	X	X		X	X	X	X	X
22	X	X	X	X	X	X	X	X
23	X	X			X	X	X	X
24				X	X		X	X
25		X	X	X	X	X	X	X
26		X	X	X	X	X	X	X
27	X	X		X		X		X
28	X	X	X	X	X	X	X	X
29						X		X
30	X	X	X	X				

de seleção de características e Decomposição Temporal forneceu resultados eficientes para os classificador J48 e 3-NN.

As tabelas 4.7 e 4.8 apresentam quais foram as características selecionadas para cada classificador e cada segmento. O símbolo # indica o número da característica.

4.4 Decomposição Temporal-Espacial com Seleção de Características

Os objetivos deste quarto experimento foram:

- Verificar se a estratégia de Decomposição Temporal–Espacial pode se beneficiar do uso do método de seleção de atributos.
- Verificar se a estratégia de Decomposição Espacial com o método de seleção de atributos aumenta o desempenho dos classificadores treinados utilizando a música inteira.
- Comparar o desempenho dos classificadores utilizando seleção de atributos com Decomposição Temporal–Espacial em relação aos métodos anteriores e também com a música inteira.

Nesta seção visando uma avaliação geral dos resultados e uma comparação com os experimentos anteriores, em todas as tabelas os resultados obtidos combinando o método de seleção de atributos com a combinação de estratégia de decomposição do espaço do problema são apresentados nas colunas FS-OAA para o método utilizando a estratégia de OAA e FS-RR para o método utilizando a estratégia de RR. As demais colunas são referentes aos experimentos anteriores onde: BL indica o *baseline* que foi definido no experimento da seção 4.1, ou seja, sem nenhum método de decomposição do espaço do problema ou seleção de atributos; OAA e RR indicam os resultados obtidos no experimento da seção 4.2, utilizando as estratégia de decomposição do espaço do problema; FS indicam os resultados obtidos no experimento da seção 4.3 utilizando apenas o método de seleção de atributos. Os resultados obtidos para os segmentos Seg_{beg}, Seg_{mid} e Seg_{end} são apresentados nas tabelas 4.9, 4.10 e 4.11 respectivamente.

A análise dos resultados apresentados na tabela 4.9 indica que para os classificadores J48 e 3-NN o método de seleção de características com Decomposição Temporal–Espacial fornece melhores resultados considerando o Seg_{beg}. Para a rede MLP o método produz melhores resultados do que utilizando apenas seleção de características. Porém,

Tabela 4.9: Taxa de Reconhecimento (%) utilizando as diversas estratégias no Seg_{beg}.

	Seg_{beg}					
Classificador	BL	OAA	RR	FS	FS-OAA	FS-RR
J48	39.60	41.56	45.96	44.70	43.52	**48.53**
3-NN	45.83	45.83	45.83	51.19	51.73	**53.36**
MLP	53.96	52.53	**55.06**	52.73	53.99	54.13
NB	44.43	42.76	44.43	**45.43**	43.46	45.39
SVM	–	23.63	**57.43**	–	26.16	57.13

ainda é inferior aos resultados obtidos utilizando apenas Decomposição Temporal–Espacial sem FS com a estratégia de decomposição RR. Para o classificador NB os resultados são melhores do que o BL e as estratégias anteriores, mas ainda ainda são inferiores ao método FS. Para o SVM os resultados são inferiores aos obtidos utilizando Decomposição Temporal–Espacial com RR.

Tabela 4.10: Taxa de reconhecimento (%) utilizando as diversas estratégias no Seg_{mid}.

	Seg_{mid}					
Classificador	BL	OAA	RR	FS	FS-OAA	FS-RR
J48	44.44	44.56	49.93	45.76	45.09	**50.86**
3-NN	56.26	56.26	56.26	60.02	60.95	**62.55**
MLP	**56.40**	53.08	54.59	54.73	54.76	49.76
NB	47.76	45.83	47.79	50.09	48.79	**50.69**
SVM	–	38.62	**63.50**	–	32.86	59.70

A análise dos resultados apresentados na tabela 4.10 indica que para o Seg_{mid} o método de FS com Decomposição Temporal–Espacial com RR fornecem melhores resultados para os classificadores J48, 3-NN e NB. Para a rede MLP os resultados obtidos ainda são inferiores ao BL. Para o SVM os resultados forem inferiores aos anteriores em ambos os casos.

Tabela 4.11: Taxa de reconhecimento (%) utilizando as diversas estratégias no Seg_{end}.

	Seg_{end}					
Classificador	BL	OAA	RR	FS	FS-OAA	FS-RR
J48	38.80	38.42	45.53	38.73	38.99	**45.86**
3-NN	48.43	48.43	48.43	51.11	51.10	**53.49**
MLP	48.26	**51.96**	51.92	47.86	50.53	49.64
NB	39.13	37.26	39.19	**39.66**	37.63	39.59
SVM	–	28.89	54.60	–	28.22	**55.33**

A análise dos resultados apresentados na tabela 4.11 indica que para o Seg_{end} os resutados só foram superiores para o classificador 3-NN. Nos demais os resultados foram similares, utilizando RR, para os classificadores SVM, NB, e J48. Para a rede MLP não houve melhora, sendo que os métodos de OAA e RR sem FS fornecem melhores resultados.

Tabela 4.12: Taxa de reconhecimento (%) utilizando as diversas estratégias na música inteira (MI).

	MI					
Classificador	BL	OAA	RR	FS	FS-OAA	FS-RR
J48	44.20	43.79	**50.63**	45.13	45.09	49.39
3-NN	57.96	57.96	**59.93**	57.30	58.22	59.47
MLP	56.46	**58.76**	57.86	53.56	53.82	54.46
NB	48.00	45.96	48.16	**49.63**	47.83	**49.63**
SVM	–	37.46	**63.40**	–	32.92	61.13

A análise dos resultados apresentados na tabela 4.12 indica que para o classificador J48, os métodos de FS-OAA e FS-RR não tiveram melhoras significativas, a melhor taxa de acerto foi obtida com o método RR. Para o 3-NN os resultados foram similares, porém a melhor taxa de acerto foi obtida com o método de FS sem RR. Para o MLP e SVM os resultados foram inferiores aos obtidos nos experimentos anteriores. Somente para o classificador NB com o FS-RR é que os resultados foram melhores do que os anteriores.

Tabela 4.13: Taxa de Reconhecimento (%) utilizando as técnicas de combinação.

	Ensembles					
Classificador	BL	OAA	RR	FS	FS-OAA	FS-RR
J48	47.33	49.63	54.06	50.10	50.03	**55.46**
3-NN	60.46	59.96	61.12	63.20	62.77	**64.10**
MLP	59.43	**61.03**	59.79	59.30	60.96	56.86
NB	46.03	43.43	47.19	47.10	44.96	**49.79**
SVM	–	30.79	**65.06**	–	29.47	63.03

A análise dos resultados apresentados na tabela 4.12 indica que para os classificadores J48, 3-NN e NB o método FS-RR fornece melhora nos resultados. Para a rede MLP o método FS-OAA produz resultados similares ao do OAA sem seleção de atributos. Para o SVM os melhores resultados são obtidos sem o uso de FS.

4.5 Avaliação e Discussão dos Resultados

Nesta seção é realizada a avaliação e discussão dos resultados obtidos no desenvolvimento deste trabalho.

No 1º experimento foi possível observar que em todos os casos utilizar o Seg_{mid} fornece melhores resultados do que utilizar a abordagem comum na literatura que utiliza o Seg_{beg}. Essa observação já havia sido feita em um experimento anterior (SILLA JR.; KAESTNER; KOERICH, 2006a) utilizando uma versão da *Latin Music Database* com apenas dois gêneros. Além disto, foi possível verificar que o uso das técnicas de combinação sempre fornece resultados melhores do que utilizar apenas um único segmento da música e também sempre produz resultados melhores do que utilizar a música inteira no caso dos classificadores mais robustos (SVM, MLP e 3-NN). Isto mostra que a questão levantada no trabalho de (MCKAY; FUJINAGA, 2006) onde era argumentado que utilizar as médias das características ao longo da música inteira poderia ser uma abordagem limitada, de fato é verificada.

No 2º experimento o método de Decomposição Temporal–Espacial sempre apresenta para todos os classificadores uma melhora na taxa de reconhecimento utilizando a estratégia de *Round Robin*. Apenas para a rede MLP é que uma melhora na taxa de acerto é obtida utilizando a estratégia de Um-Contra-Todos (*One-Against-All*). Utilizando o MLP com OAA de exemplo é possível observar como apesar de no caso dos segmentos individuais o desempenho do *ensemble* melhorou, isto se deve a um aumento na diversidade introduzida pelo método de decomposição do problema que levou a uma melhora no uso da estratégia de combinação. Outra observação importante é que, em relação a um experimento anterior (SILLA JR.; KAESTNER; KOERICH, 2006b) utilizando uma versão da base com 4 gêneros o método de Decomposição Temporal–Espacial, os resultados obtidos utilizando o método na base de 10 gêneros são superiores aos resultados utilizando os segmentos individualmente. Isto confirma as expectativas apresentadas no trabalho citado de que o método de Decomposição Temporal–Espacial fornece melhores resultados com um número maior de classes, visto que as funções de decisão e separação das classes se tornam mais complexas.

No 3º experimento o método de seleção de atributos com algoritmos genéticos para classificação dos segmentos da música se mostrou eficiente para os classificadores J48, 3-NN e NB e ineficiente para SVM e MLP. Para a música inteira o método de seleção de atributos somente forneceu resultados melhores para J48 e NB. O uso da Decomposição Temporal em conjunto com o método de seleção de atributos forneceu uma melhora significativa para o classificador 3-NN. Apesar desta melhora ter sido apenas para este

classificador, este resultado corrobora com os experimentos recentemente realizados em (YASLAN; CATALTEPE, 2006) onde os resultados positivos também foram obtidos utilizando o classificador k-NN.

No 4º experimento foi integrado o método de Decomposição Temporal–Espacial ao módulo de seleção de atributos e os resultados obtidos foram comparados com os resultados realizados nos três experimentos anteriores. Para a classificação dos segmentos individuais o método apresenta melhora para Seg_{beg} e Seg_{mid} com os classificadores J48, 3-NN e NB. Para Seg_{end} um aumento na taxa de reconhecimento é obtida com os classificadores J48, 3-NN e SVM. Em todos os casos os resultados são obtidos com a estratégia de RR. Para a música inteira somente o classificador NB apresenta resultados melhores em relação aos demais experimentos. Para a estratégia de combinação é obtida uma melhora nos resultados dos classificadores J48, 3-NN e NB.

No intuito de verificar qual a melhor estratégia para cada classificador utilizado foi criada a tabela 4.14 que apresenta uma comparação das melhoras taxas de classificação considerando todos os experimentos realizados neste trabalho, onde % indica a taxa de classificação e M indica o método utilizado.

Tabela 4.14: Comparação das melhores taxas de classificação (%) em cada segmento, na música inteira e na combinação

	Seg_{beg}		Seg_{mid}		Seg_{end}		MI		*Ensemble*	
	%	M	%	M	%	M	%	M	%	M
J48	48.53	FS-RR	50.86	FS-RR	45.86	FS-RR	50.63	RR	**55.46**	FS-RR
3-NN	53.36	FS-RR	62.55	FS-RR	53.49	FS-RR	59.93	RR	**64.86**	FS-WS II
MLP	55.06	RR	56.40	BL	51.96	OAA	58.76	OAA	**62.50**	BL-PROD
NB	45.43	RR	50.69	FS-RR	39.66	FS	49.63	FS	**51.46**	FS-WS I
SVM	57.43	RR	63.50	RR	55.33	FS-RR	64.40	RR	**66.76**	RR-WS I

A análise dos resultados apresentados na tabela 4.14 mostra que em todos os casos os melhores resultados são obtidos utilizando alguma das técnicas de combinação de classificadores.

Capítulo 5

Considerações Finais

Neste trabalho foi apresentada uma abordagem baseada em combinação de classificadores e seleção de características para a tarefa de classificação automática de gêneros musicais. Para realizar a tarefa de classificação foram utilizados algoritmos clássicos de aprendizagem supervisionada como Árvores de Decisão (J48), Naïve Bayes (NB), k vizinhos mais próximos (k-NN), Máquinas de Suporte Vetorial (SVM), e redes neurais MLP. Para realizar os experimentos foi construída uma base inédita para a tarefa denominada Latin Music Database que contém 3.160 músicas de 10 gêneros. Após a construção da base foram realizados quatro experimentos de classificação automática de gêneros musicais no intuito de verificar: (1) se o método de Decomposição Temporal fornece resultados melhores do que utilizar os segmentos individuais ou a música inteira; (2) se o método de Decomposição Temporal–Espacial fornece resultados superiores aos obtidos pela Decomposição Temporal; (3) se o uso de um mecanismo de seleção de atributos baseado em algoritmos genéticos melhora a taxa de reconhecimento dos gêneros; e (4) se o uso desse mecanismo de seleção poderia melhorar os resultados dos métodos anteriores.

Cada um destes experimentos foi realizado e analisado, concluindo-se que as estratégias baseadas na combinação de classificadores sempre fornecem melhores resultados relativamente ao uso de segmentos isolados ou características extraídas da música inteira. Durante a realização diversas contribuições científicas originais, algumas limitações e trabalhos futuros foram identificados.

As principais contribuições deste trabalho para o estado da arte da classificação automática de gêneros musicais e para a recuperação de informações musicais, são:

- A construção de uma nova base de dados intitulada de *Latin Music Database*, que utilizou um procedimento inédito para a atribuição de gêneros as músicas e pode ser utilizada para diversas tarefas relacionadas a MIR, não limitando-se a classificação automática de gêneros musicais;

- A verificação de que o uso de combinação de classificadores utilizando classificadores treinados em diferentes trechos da música é uma forma eficiente de realizar a tarefa, obtendo resultados superiores ao uso de características extraídas da música inteira e também dos respectivos segmentos individuais. Uma observação importante é que em nenhum trabalho da literatura é verificada a questão de que se utilizar um trecho da música (usualmente o início) é melhor ou pior do que utilizar a música inteira, tornado essa avaliação pioneira neste sentido;

- A avaliação do uso de mecanismos de seleção de atributos baseados em algoritmos genéticos, visto que este tópico tem sido somente indicado como trabalho futuro em publicações recentes da área;

- Uma contribuição menor porém pertinente é que ficou evidente que se for utilizado apenas um segmento da música o ideal é utilizar o Seg_{mid} ao invés do Seg_{beg} como é feito na literatura;

- Este é o primeiro trabalho na área a considerar efetivamente a variação temporal da música, pois a decisão, no caso da combinação de classificadores leva em conta características extraídas de diferentes segmentos da música.

A principal limitação desse trabalho está no módulo de seleção de características que é dependente do MARSYAS. Porém, pela forma como o sistema de classificação foi implementado, ele pode ser substituído facilmente por outros softwares de extração de características como o JAudio sem nenhuma dificuldade. Outra limitação deste trabalho é que não foram realizados experimentos no intuito de verificar a quantidade de segmentos e/ou a duração de cada segmento para obter uma melhor taxa de classificação. Foi definido no início do projeto que seriam utilizados três segmentos de trinta segundos e esta decisão foi seguida em todos os experimentos.

Com o desenvolvimento da *Latin Music Database*, existe um grande número de possibilidades para a continuação deste trabalho. Primeiramente pretende-se integrar a base ao sistema OMEN de forma a disponibilizá-la para todos os pesquisadores da comunidade MIR sem ter problemas com as questões de distribuição em função dos direitos autorais das músicas. Além disto, seria interessante repetir os experimentos realizados na base de dados CODAICH que logo estará integrada ao OMEN.

Também poderiam ser utilizadas as novas ferramentas disponíveis como o JAudio para extração de outros vetores de características. Outra opção seria estudar o uso de métodos de *stacking* como foi realizado no trabalho de (YASLAN; CATALTEPE, 2006) para verificar se isso poderia trazer uma melhora para a tarefa de classificação.

Uma outra possibilidade interessante seria utilizar a combinação de classificadores utilizando ténicas de *webmining* em conjunto com as características extraídas do conteúdo do sinal de áudio, além de um estudo comparativo utilizando diversas técnicas de seleção de atributos, dentre elas o uso de algoritmos genéticos multi-objetivos e análise de componentes principais.

Referências Bibliográficas

AUCOUTURIER, J. J.; PACHET, F. Representing musical genre: A state of the art. *Journal of New Music Research*, v. 32, n. 1, p. 83–93, 2003.

BLUM, A.; LANGLEY, P. Selection of relevant features and examples in machine learning. *Artificial Intelligence*, v. 97, n. 1-2, p. 245–271, 1997.

BRAY, S.; TZANETAKIS, G. Distributed audio feature extraction for music. In: INTERNATIONAL CONFERENCE ON MUSIC INFORMATION RETRIEVAL, 6., 2005, London, UK. *Anais...* London, UK: ISMIR, 2005. p. 434–437.

COSTA, C. H. L.; VALLE JR., J. D.; KOERICH, A. L. Automatic classification of audio data. In: IEEE INTERNATIONAL CONFERENCE ON SYSTEMS, MAN, AND CYBERNETICS, 2004, Haia, Holanda. *Anais...* Haia, Holanda: IEEE Press, 2004. p. 562–567.

DASH, M.; LIU, H. Feature selection for classification. *Intelligent Data Analysis*, v. 1, n. 1–4, p. 131–156, 1997.

DEMSAR, J. Statistical comparisons of classifiers over multiple data sets. *Journal of Machine Learning Research*, v. 7, p. 1–30, 2006.

DIETTERICH, T. G. Ensemble methods in machine learning. In: INTERNATIONAL WORKSHOP ON MULTIPLE CLASSIFIER SYSTEM, 1., 2000, Cagliari, Italy. *Anais...* Cagliari, Italy: Springer, 2000. (Lecture Notes in Computer Science, 1857), p. 1–15.

DOWNIE, J. S.; CUNNINGHAM, S. J. Toward a theory of music information retrieval queries: System design implications. In: INTERNATIONAL CONFERENCE ON MUSIC INFORMATION RETRIEVAL, 3., 2002, Paris, France. *Anais...* Paris, France: ISMIR, 2002. p. 299–300.

FABBRI, F. *Browsing Music Spaces: Categories and the Musical Mind*. 1999. Disponível em: <http://www.mediamusicstudies.net/tagg/others/ffabbri9907.html>. Acesso em: 20 jan. 2007.

FIEBRINK, R.; FUJINAGA, I. Feature selection pitfalls and music classification. In: INTERNATIONAL CONFERENCE ON MUSIC INFORMATION RETRIEVAL, 7., 2006, Victoria, Canada. *Anais...* Victoria, Canada: ISMIR, 2006. p. 340–341.

FINGERHUT, M. The ircam multimedia library: A digital music library. In: IEEE FORUM ON RESEARCH AND TECHNOLOGY ADVANCES IN DIGITAL LIBRARIES, 1., 1999, Baltimore, USA. *Anais...* Baltimore, USA: IEEE Press, 1999. p. 19–21.

FÜRNKRANZ, J. Pairwise classification as an ensemble technique. In: EUROPEAN CONFERENCE ON MACHINE LEARNING, 13., 2002, Helsinki, Finland. *Anais...* Helsinki, Finland: Springer-Verlag, 2002. (Lecture Notes in Computer Science, v. 2430), p. 97–110.

GRIMALDI, M.; CUNNINGHAM, P.; KOKARAM, A. *An Evaluation of Alternative Feature Selection Strategies and Ensemble Techniques for Classifying Music*. 2003. Disponível em: <http://www.mee.tcd.ie/˜ moumir/articles/Marco_ECML2003-MIW.pdf>. Acesso em: 10 out. 2005.

GRIMALDI, M.; CUNNINGHAM, P.; KOKARAM, A. A wavelet packet representation of audio signals for music genre classification using different ensemble and feature selection techniques. In: ACM SIGMM INTERNATIONAL WORKSHOP ON MULTIMEDIA INFORMATION RETRIEVAL, 5., 2003, Berkeley, California. *Anais...* Berkeley, California: ACM Press, 2003. p. 102–108.

GUO, G.; LI, S. Z. Content–based audio classification and retrieval by support vector machines. *IEEE Transactions on Neural Networks*, v. 14, n. 1, p. 209–215, 2003.

HO, T. K. Nearest neighbors in random subspaces. In: ADVANCES IN PATTERN RECOGNITION, JOINT IAPR INTERNATIONAL WORKSHOPS SSPR E SPR, 1998, Sydney, Australia. *Anais...* Sydney, Australia: Springer Verlag, 1998. (Lecture Notes in Computer Science, v. 1451), p. 640–648.

HOMBURG, H. et al. A benchmark dataset for audio classification and clustering. In: INTERNATIONAL CONFERENCE ON MUSIC INFORMATION RETRIEVAL, 6., 2005, London, UK. *Anais...* London, UK: ISMIR, 2005. p. 528–531.

HU, X. et al. Mining music reviews: Promising preliminary results. In: INTERNATIONAL CONFERENCE ON MUSIC INFORMATION RETRIEVAL, 6., 2005, London, UK. *Anais...* London, UK: ISMIR, 2005. p. 536–539.

KITTLER, J. et al. On combining classifiers. *IEEE Transaction on Pattern Analysis and Machine Intelligence*, v. 20, n. 3, p. 226–239, March 1998.

KOERICH, A. L.; POITEVIN, C. Combination of homogeneous classifiers for musical genre classification. In: IEEE INTERNATIONAL CONFERENCE ON SYSTEMS, MAN, AND CYBERNETICS, 2005, Hawaii, USA. *Anais...* Hawaii, USA: IEEE Press, 2005. p. 554–559.

KOSINA, K. *Music Genre Recognition*. Dissertação (Mestrado) — Fachlochschul Hagenberg, Fachlochschul Hagenberg, 2002.

KUNCHEVA, L. I. *Combining Pattern Classifiers*. New Jersey: Wiley-Interscience, 2004. 360 p.

LEE, J. H.; DOWNIE, J. S. Survey of music information needs, uses, and seeking behaviours: preliminary findings. In: INTERNATIONAL CONFERENCE ON MUSIC INFORMATION RETRIEVAL, 5., 2004, Barcelona, Spain. *Anais...* Barcelona, Spain: ISMIR, 2004. p. 441–446.

LI, T.; OGIHARA, M. Music genre classification with taxonomy. In: IEEE INTERNATIONAL CONFERENCE ON ACOUSTICS, SPEECH,AND SIGNAL PROCESSING, 2005, Philadelphia, USA. *Anais...* Philadelphia, USA: ISMIR, 2005. p. 197–200.

LI, T.; OGIHARA, M.; LI, Q. A comparative study on content-based music genre classification. In: INTERNATIONAL ACM SIGIR CONFERENCE ON RESEARCH AND DEVELOPMENT IN INFORMAION RETRIEVAL, 26., 2003, Toronto, Canada. *Anais...* Toronto, Canada: ACM Press, 2003. p. 282–289.

LYMAN, P.; VARIAN, H. R. *How much information*. 2003. Disponível em: <http://www.sims.berkeley.edu/how-much-info-2003>. Acesso em: 25 mar. 2005.

MCENNIS, D.; MCKAY, C.; FUJINAGA, I. Overview of on-demand metadata extraction network (omen). In: INTERNATIONAL CONFERENCE ON MUSIC INFORMATION RETRIEVAL, 7., 2006, Victoria, Canada. *Anais...* Victoria, Canada: ISMIR, 2006. p. 7–12.

MCENNIS, D. et al. Jaudio: A feature extraction library. In: INTERNATIONAL CONFERENCE ON MUSIC INFORMATION RETRIEVAL, 6., 2005, London, UK. *Anais...* London, UK: ISMIR, 2005. p. 600–603.

MCKAY, C. et al. Ace: A framework for optimizing music classification. In: INTERNATIONAL CONFERENCE ON MUSIC INFORMATION RETRIEVAL, 6., 2005, London, UK. *Anais...* London, UK: ISMIR, 2005. p. 42–49.

MCKAY, C.; FUJINAGA, I. Musical genre classification: Is it worth pursuing and how can it be. In: INTERNATIONAL CONFERENCE ON MUSIC INFORMATION RETRIEVAL, 7., 2006, Victoria, Canada. *Anais...* Victoria, Canada: ISMIR, 2006. p. 101–106.

MCKAY, C.; MCENNIS, D.; FUJINAGA, I. A large publicly accessible database of annotated audio for music research. In: INTERNATIONAL CONFERENCE ON MUSIC INFORMATION RETRIEVAL, 7., 2006, Victoria, Canada. *Anais...* Victoria, Canada: ISMIR, 2006. p. 160–163.

MENG, A.; AHRENDT, P.; LARSEN, J. Improving music genre classification by short-time feature integration. In: IEEE INTERNATIONAL CONFERENCE ON ACOUSTICS, SPEECH, AND SIGNAL PROCESSING, 30., 2005, Philadelphia, USA. *Anais...* Philadelphia, PA, USA: IEEE PRESS, 2005. p. 497–500.

MOLINA, L. C.; BELANCHE, L.; NEBOT, À. Feature selection algorithms: A survey and experimental evaluation. In: IEEE INTERNATIONAL CONFERENCE ON DATA MINING, 2002, Maebashi City, Japan. *Anais...* Maebashi City, Japan: IEEE Press, 2002. p. 306–313.

PAMPALK, E.; RAUBER, A.; MERKL, D. Content–based organization and visualization of music archives. In: ACM INTERNATIONAL CONFERENCE ON MULTIMEDIA, 10., 2002, Juan-les-Pins, France. *Anais...* Juan-les-Pins, France: ACM, 2002. p. 570–579.

SHAO, X.; XU, C.; KANKANHALLI, M. S. Applying neural network on the content–based audio classification. In: INTERNATIONAL CONFERENCE ON INFORMATION, COMMUNICATIONS ANDSIGNAL PROCESSING, 4., 2003, Singapore. *Anais...* Singapore: IEEE Press, 2003. v. 3, p. 1821–1825.

SILLA JR., C. N.; KAESTNER, C. A. A.; KOERICH, A. L. Classifica¸ção automática de gêneros musicais utilizando métodos de bagging e boosting. In: BRAZILIAN

SYMPOSIUM ON COMPUTER MUSIC, 10., 2005, Belo Horizonte, Brasil. *Anais...* Belo Horizonte: SBC, 2005. p. 48–57.

SILLA JR., C. N.; KAESTNER, C. A. A.; KOERICH, A. L. Automatic genre classification of latin music using ensemble of classifiers. In: SEMINÁRIO INTEGRADO DE SOFTWARE E HARDWARE, 33., 2006, Campo Grande, MS, Brasil. *Anais...* São Paulo: SONOPRESS, 2006. p. 47–53.

SILLA JR., C. N.; KAESTNER, C. A. A.; KOERICH, A. L. Time-space ensemble strategies for automatic music genre classification. In: SICHMAN, J. S.; COELHO, H.; REZENDE, S. O. (Ed.). *Advances in Artificial Intelligence - IBERAMIA-SBIA 2006*. Ribeirão Preto, Brasil: Springer, 2006. (Lecture Notes in Artificial Intelligence, v. 4140), p. 339–348.

SWELDENS, W.; PIESSENS, R. Wavelet sampling techniques. In: STATISTICAL COMPUTING SECTION, 1993, USA. *Anais...* USA: American Statistical Association, 1993. p. 20–29.

TZANETAKIS, G.; COOK, P. Marsyas: A framework for audio analysis. *Organized Sound*, v. 4, n. 3, p. 169–175, 1999.

TZANETAKIS, G.; COOK, P. Musical genre classification of audio signals. *IEEE Transactions on Speech and Audio Processing*, v. 10, n. 5, p. 293–302, 2002.

VIGNOLI, F. Digital music interaction concepts: a user study. In: INTERNATIONAL CONFERENCE ON MUSIC INFORMATION RETRIEVAL, 5., 2004, Barcelona, Spain. *Anais...* Barcelona, Spain: ISMIR, 2004.

WEST, K.; COX, S. Finding an optimal segmentation for audio genre classification. In: INTERNATIONAL CONFERENCE ON MUSIC INFORMATION RETRIEVAL, 6., 2005, London, UK. *Anais...* London, UK: ISMIR, 2005. p. 680–685.

WITTEN, I. H.; FRANK, E. *Data Mining: Practical machine learning tools and techniques*. 2nd. ed. San Francisco: Morgan Kaufmann, 2005.

YANG, J.; HONAVAR, V. Feature subset selection using a genetic algorithm. *IEEE Intelligent Systems*, v. 13, n. 2, p. 44–49, 1998.

YASLAN, Y.; CATALTEPE, Z. Audio music genre classification using different classifiers and feature selection methods. In: INTERNATIONAL CONFERENCE ON PATTERN RECOGNITION, 18., 2006, Hong Kong. *Anais...* Hong Kong: ICPR, 2006. p. 573–576.

ZHANG, T.; KUO, C. C. J. Audio content analysis for online audiovisual data segmentation and classification. *IEEE Transactions on Speech and Audio Processing*, v. 9, n. 4, p. 441–457, 2001.

Glossário

Algoritmo Genético (AG)

É uma técnica de procura utilizada na ciência da computação para achar soluções aproximadas em problemas de otimização e busca. Algoritmos genéticos são uma classe particular de algoritmos evolutivos que usam técnicas inspiradas pela biologia evolutiva como hereditariedade, mutação, seleção natural e recombinação.

Altura da nota (*Pitch*)

Refere-se à forma como o ouvido humano percebe a freqüência dos sons. As baixas freqüências são percebidas como sons graves e as mais altas como sons agudos.

Aprendizagem Supervisionada

É um dos tipos de algoritmo de aprendizagem de máquina onde: na fase de treinamento são apresentados os objetos e suas respectivas classes. No contexto deste trabalho os objetos são as músicas e as classes os gêneros musicais.

Classificador

É um sistema que faz a classificação de objetos. Em linhas gerais, a tarefa do classificador consiste em organizar os objetos de um dado universo em grupos ou categorias, com um propósito específico.

Espectro Sonoro (*Timbral Texture*)

É a distribuição, no domínio das freqüências, do conjunto de todas as ondas que formam um som. O espectro é composto das amplitudes da freqüência fundamental e de cada um dos harmônicos ou parciais que soam junto à fundamental. O espectro de um som define sua forma de onda e é um dos componentes do timbre de uma voz ou instrumento musical.

Gênero Musical

É utilizada a definição de (FABBRI, 1999): "Um tipo de música, como ela é aceita por uma comunidade por qualquer razão, propósito ou critério".

Meta-Informações Textuais

São informações a respeito das músicas que estão sendo utilizadas como título e artista.

Padrão Rítmico (*Beat*)

Refere-se à forma como o ouvido humano percebe a batida de uma música.

Anexo A

O Formato do Rótulo ID3

O formato de áudio MPEG layer I, layer II e layer III (MP3) não tem uma forma nativa de armazenar informação sobre o seu conteúdo, exceto por algums parametros simples do tipo sim/não como “privado”, “copyrighted”e “original home”(indicando que esse arquivo é o original e não uma cópia). Uma solução para esse problema apresentada no programa “Studio3”, desenvolvido por Eric Kemp em 1996, consiste em adicionar uma pequena quantidade de informações extras no final de um arquivo MP3 para poder armazenar informações a respeito do conteúdo daquele arquivo.

A localização do rótulo ID3, como essa informação ficou conhecida, foi provavelmente escolhida, pois havia pouca chance de apresentar problemas para os decodificadores. No intuito de fazer o rótulo ser facilmente detectado um tamanho fixo de 128 bytes foi escolhido. A Figura A.1 apresenta o layout do rótulo ID3, enquanto a Tabela A.1 apresenta o counteúdo das informações do róulo.

Tabela A.1: Informações do conteúdo do rótulo ID3

Informação	Tamanho
Título	30 caracteres
Artista	30 caracteres
Album	30 caracteres
Ano	4 caracteres
Comentário	30 caracteres
Gênero	1 byte

Como pode ser visto na Tabela A.1 o campo gênero possui apenas um byte, e além disso o padrão original só continha uma lista fixa de 80 gêneros (apresentados na tabela A.2) que foi criada sem critério algum. Como o campo Gênero é representado por um byte, como visto na Tabela A.1, outros desenvolvedores adicionaram suas próprias listas de gêneros a lista original, impossibilitando uma padronização.

Figura A.1: Esquema do Formato do Rótulo ID3 (www.id3.org)

Tabela A.2: Mapeamento dos Gêneros no Padrão ID3

#	Gênero	#	Gênero	#	Gênero	#	Gênero
0	Blues	20	Alternative	40	AlternRock	60	Top 40
1	Classic Rock	21	Ska	41	Bass	61	Christian Rap
2	Country	22	Death Metal	42	Soul	62	Pop/Funk
3	Dance	23	Pranks	43	Punk	63	Jungle
4	Disco	24	Soundtrack	44	Space	64	Native American
5	Funk	25	Euro-Techno	45	Meditative	65	Cabaret
6	Grunge	26	Ambient	46	Instrumental Pop	66	New Wave
7	Hip-Hop	27	Trip-Hop	47	Instrumental Rock	67	Psychadelic
8	Jazz	28	Vocal	48	Ethnic	68	Rave
9	Metal	29	Jazz+Funk	49	Gothic	69	Showtunes
10	New Age	30	Fusion	50	Darkwave	70	Trailer
11	Oldies	31	Trance	51	Techno-Industrial	71	Lo-Fi
12	Other	32	Classical	52	Electronic	72	Tribal
13	Pop	33	Instrumental	53	Pop-Folk	73	Acid Punk
14	R&B	34	Acid	54	Eurodance	74	Acid Jazz
15	Rap	35	House	55	Dream	75	Polka
16	Reggae	36	Game	56	Southern Rock	76	Retro
17	Rock	37	Sound Clip	57	Comedy	77	Musical
18	Techno	38	Gospel	58	Cult	78	Rock & Roll
19	Industrial	39	Noise	59	Gangsta	79	Hard Rock