Classificação Automática de Gêneros Musicais Utilizando Métodos de *Bagging* e *Boosting*

Carlos N. Silla Jr. , Celso A. A. Kaestner , Alessandro L. Koerich

Pontifícia Universidade Católica do Paraná (PUC-PR)
Programa de Pós-Graduação em Informática Aplicada (PPGIA)
Rua Imaculada Conceição 1155, 80215-901
Curitiba - PR - Brasil

{silla,kaestner,alekoe}@ppgia.pucpr.br

Abstract. *This paper presents a study that uses meta-learning techniques to the task of automatic music genre classification. The meta-learning techniques we used are Bagging and Boosting. In both cases the component classifiers used in both approaches are Decision Trees, k-NN (k nearest neighbors) and Naive Bayes. The experiments were performed on a dataset containing 1.000 songs with 10 different genres. The achieved results show that the Bagging approach is promising while the Boosting approach seems to be inadequate to the problem.*

Resumo. *Este artigo apresenta um estudo utilizando técnicas de meta-aprendizagem para o problema de classificação automática de gêneros musicais. As técnicas de meta-aprendizagem utilizadas foram Bagging e Boosting. Em ambos os casos usando como classificadores componentes os algoritmos de Árvores de Decisão, k-NN (k vizinhos mais próximos) e Naive Bayes. Os experimentos foram realizados utilizando uma base contendo 1.000 músicas de 10 gêneros diferentes. Os resultados obtidos com o algoritmo de Bagging foram positivos, enquanto o uso do método de Boosting apresentou resultados aquém dos esperados.*

1. Introdução

A quantidade de conteúdo multimídia disponível on-line criou uma necessidade de ferramentas capazes de organizar e gerenciar essa grande quantidade de informações [Fingerhut, 1999] [Pampalk et al., 2002]. No momento, a maior parte das informações sobre dados multimídia são classificadas e organizadas baseadas em meta-informações textuais que são associadas ao seu conteúdo, como é o caso dos tags ID3 nos arquivos de áudio com o formato .mp3. Apesar destas informações serem relevantes para as tarefas de indexação, busca e recuperação, elas são geradas manualmente e então associados com o arquivo multimídia.

A música digital é um dos mais importantes tipos de dados distribuídos na Internet. Como organizar e processar essa grande variedade e quantidade de dados de maneira eficiente para permitir indexar, buscar e recuperar é um grande desafio [Foote, 1999] [Guo and Li, 2003]. Existem muitos estudos e métodos sobre a análise de conteúdo de áudio usando diferentes características e métodos [Pampalk et al., 2002], [Guo and Li, 2003], [Zhang and Kuo, 2001], [Liu et al., 2003], [Aucouturier and Pachet, 2003].

Apesar dos esforços de vários pesquisadores, a classificação automática de áudio é realizada com relativa precisão somente em problemas simples, como diferenciar fala

e música [Carey et al., 1999]. Poucos trabalhos tratam com a tarefa de classificação automática de gêneros musicais [Pye, 2000], [Kosina, 2002], [Tzanetakis and Cook, 2002], [Liu et al., 2003], [Aucouturier and Pachet, 2003]. Um gênero musical é uma descrição importante que tem sido utilizado para classificar e caracterizar músicas digitais e para organizar grandes coleções disponíveis na Web [Tzanetakis and Cook, 2002] [Shao et al., 2003] [Liu et al., 2003], [Aucouturier and Pachet, 2003]. Os gêneros musicais são rótulos categóricos criados por especialistas humanos assim como por amadores para determinar títulos de músicas. Esses rótulos são relacionados com a instrumentalização utilizada, estrutura rítmica e conteúdo harmônico da música. Entretanto, um gênero musical é um conceito relativamente subjetivo, e mesmo a indústria musical muitas vezes é contraditória ao atribuir gêneros musicais para as músicas. Uma prática muito como é que as músicas são categorizadas de acordo com o perfil do artista. Adicionalmente a classificação de músicas tem sido desenvolvida normalmente para álbuns, e não é aplicável diretamente para as faixas do álbum [Aucouturier and Pachet, 2003]. Dessa forma a classificação automática de gêneros musicais pode auxiliar ou substituir o usuário humano nesse processo, assim como prover um componente importante para um sistema de recuperação de informações para músicas.

A classificação automática de gêneros musicais é um tema relativamente novo e nos últimos anos tem atraído a atenção de diversos pesquisadores. Pye [Pye, 2000] usou os coeficientes cepstrais de freqüência-Mel (*Mel-frequency cepstral coefficients* - MFCC) e um *Gaussian mixture model* para classificar músicas em cinco classes: blues, easy listening, classic, opera, dance e rock. Tzanetakis e Cook [Tzanetakis and Cook, 2002] propuseram um conjunto de características para modelar os sinais musicais que estão relacionadas a textura timbral (*timbral texture*), ritmo (*rhythm*) e variações da freqüência da vibração (*pitch*). Nesse trabalho eles utilizaram dois algoritmos de classificação (Gaussian mixture model e k-NN) aplicando as características extraídas. Shao [Shao et al., 2003] usou um modelo de classificação não supervisionada baseada em Cadeias Escondidas de Markov (HMMs). Liu et al. [Liu et al., 2003] propôs um novo método de extrair características baseado nos histogramas computados utilizando os coeficientes da Daubechies Wavelet. Para classificação eles utilizaram diferentes algoritmos como: Support Vector Machines (SVM), k-NN, *Gaussian mixture models* e análise discriminante linear (*Linear Discrimant Analysis*). Uma breve revisão dos trabalhos anteriores na área de classificação automática de gêneros musicais mostra que o interesse atual é o desenvolvimento de novos conjuntos de características e a utilização de diferentes classificadores. Um ponto comum nos trabalhos anteriores é que eles utilizam sempre um único classificador.

Neste trabalho é proposta uma nova abordagem para o problema de classificação automática de gêneros musicais utilizando a combinação de classificadores componentes através de técnicas de meta-aprendizagem. Neste trabalho foram utilizadas as técnicas de *Bagging* [Breiman, 1996] e *Boosting* [Mitchell, 1997] com o algoritmo AdaBoostM1 [Freund and Schapire, 1996]. O restante deste trabalho está organizado da seguinte maneira: na seção 2 é apresentada uma visão geral do sistema; na seção 3 são apresentadas as características utilizadas; na seção 4 são apresentadas breves descrições dos métodos de meta-aprendizagem empregados; na seção 5 são apresentados os experimentos realizados e os resultados obtidos; e finalmente na seção 6 são apresentadas as conclusões deste trabalho.

2. Visão Geral do Sistema

O sistema de classificação de gêneros musicais utilizando técnicas de meta-aprendizagem é composto de dois módulos principais (Figura 1): um módulo para a extração de carac-

terísticas e um módulo de classificação de gêneros musicais. Inicialmente um conjunto de características é extraído a partir dos trinta primeiros segundos do sinal de áudio. Ao todo são extraídas trinta diferentes características que formarão um vetor de características.

Por se tratar de um sistema que utiliza algoritmos de classificação, o sistema opera em dois modos: treinamento e teste. No modo de treinamento os vetores de características são utilizados com seus respectivos rótulos (neste contexto o gênero da música) para treinar os algoritmos de classificação utilizados pelas técnicas de meta-aprendizagem *Bagging* e *Boosting*. Os rótulos consistem na informação textual representando o gênero musical atribuído àquela música por especialistas humanos. No modo de classificação, a música cujo gênero é desconhecido é fornecida ao sistema. Um vetor de características é gerado a partir dos trinta primeiros segundos da música e este vetor é passado ao algoritmo de classificação que atribui um gênero à música.

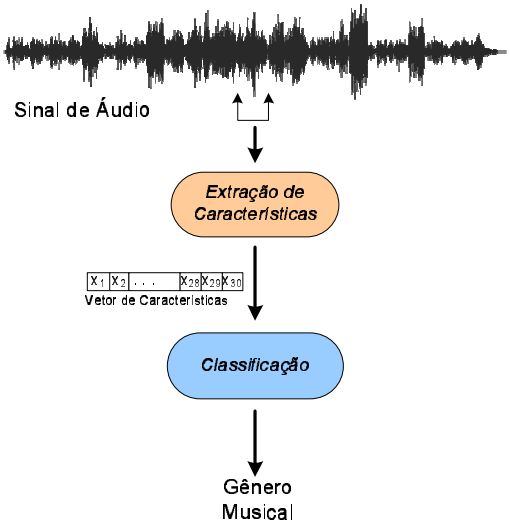

Figura 1: Visão Geral do Sistema

3. Extração de Características

O conjunto de características utilizadas neste trabalho foi originalmente proposto por [Tzanetakis and Cook, 2002] e utilizado em outros trabalhos [Kosina, 2002]. São considerados três tipos de características: textura timbral (*Timbral texture*), relacionadas à batida (*beat-related*) e relacionadas as variações da freqüência da vibração (*pitch-related*). Características de textura timbral incluem a média e a variância do centróide espectral, do *rolloff* espectral, do fluxo espectral, das taxas de cruzamento zero, coeficientes cepstrais de freqüência-Mel (MFCC), e da baixa energia. Características relacionadas à batida 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 30 dimensões (Textura Timbral: 9 FFT + 10 MFCC; Ritmo: 6; Pitch: 5) que posteriormente são utilizados no treinamento dos diferentes classificadores de maneira supervisionada. A seguir são descritas as características extraídas das músicas.

3.1. Textura Timbral

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 1.

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

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 2 na qual 85% da magnitude da distribuição está concentrada.

$$\sum_{n=1}^{R_t} M_t[n] = 0.85 \sum_{n=1}^{N} M_t[n] \tag{2}$$

Fluxo Espectral (*Spectral Flux*) é uma medida da mudança espectral local e é definido como apresentado na equação 3.

$$F_t = \sum_{n=1}^{N} (N_t[n] - N_{t-1}[n])^2 \tag{3}$$

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 4.

$$Z_t = \frac{1}{2} \sum_{n=1}^{N} |sign(x[n]) - sign(x[n-1])| \tag{4}$$

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 transformada de Fourier (*Short Time Fourier Transform* - STFT). 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 and 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. Esses 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 nessa seção, em um número de janelas de análise. O termo janela de textura é utilizado para descrever essa 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, a textura timbral 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. Características Relacionadas à Batida (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 da característica da batida tenta encontrar a batida principal da música e de seu período em BPM (batidas por minuto). Além desta, é calculada também a 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 [Sweldens and Piessens, 1993]. Após essa 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, *downsampling* e remoção das médias [Kosina, 2002, Tzanetakis and Cook, 2002].

Após o passo da extração, os envelopes de cada banda são somados e a autocorrelação resultante é calculada. Esse 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. Esse 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. Essa 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 a 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.

3.3. Características Relacionadas ao Pitch (Pitch-Related)

O conjunto de características de conteúdo pitch é baseado em múltiplas técnicas de detecção de pitch. Nesse 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 dessa 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).

4. Classificação Utilizando Métodos de Meta-Apredizagem

Com os vetores de características calculados na seção 3, é possível utilizar algoritmos padrões de aprendizado de máquina. Neste trabalho são utilizados os métodos de meta-aprendizagem de *Bagging* [Breiman, 1996] e *Boosting* [Mitchell, 1997] com os classificadores componentes de árvores de decisão [Quinlan, 1993] (versão Java J4.8), Naive Bayes [Mitchell, 1997] e k-NN [Aha et al., 1991].

O método de *Bagging* consiste em utilizar múltiplas versões de um conjunto de treinamento, cada versão é criada selecionando aleatoriamente $n' < n$ amostras do conjunto de treinamento D, com reposição. Cada um destas versões é utilizada para treinar diferentes "classificadores componentes" e a decisão da classificação final é baseada no voto de cada componente. Neste trabalho é utilizada uma abordagem tradicional, onde todos os empregam o mesmo método de classificação.

O método de *Boosting* tem como objetivo melhorar a precisão de qualquer algoritmo de aprendizagem. Para isso o procedimento utilizado é o seguinte: (1) é criado um classificador com precisão sobre o conjunto de treinamento; (2) cria-se um classificador com precisão sobre o conjunto de treinamento maior do que a média. Adicionalmente novos classificadores componentes são adicionados para formar um conjunto cuja regra de decisão tenha uma alta precisão arbitrária sobre o conjunto de treinamento. Neste trabalho foi utilizada uma variação do algoritmo de *Boosting* denominada AdaBoostM1 [Freund and Schapire, 1996].

Tabela 1: Resultados da Classificação Automática de Gêneros Musicais (base padrão)

Classificador	Individual	Pior Bag	Melhor Bag	Pior Boost	Melhor Boost
J4.8	52.1	57.4	66.9	59.9	68.1
k-NN (k=1)	49.3	47.1	52.8	49.3	49.5
k-NN (k=2)	45.4	48.8	54.3	46.4	47.2
k-NN (k=3)	49.9	50.2	55.0	46.3	46.3
k-NN (k=4)	49.8	51.0	54.7	49.8	49.8
k-NN (k=5)	50.2	50.4	54.7	50.2	50.2
Naive Bayes	57.0	57.2	59.4	56.6	56.6

Tabela 2: Resultados da Classificação Automática de Gêneros Musicais (base aleatória)

Classificador	Individual	Pior Bag	Melhor Bag	Pior Boost	Melhor Boost
J4.8	50.1	57.9	67.5	58.6	65.9
k-NN (k=1)	52.0	52.1	56.5	52.0	52.1
k-NN (k=2)	46.7	50.9	56.8	47.8	48.2
k-NN (k=3)	51.8	51.4	56.5	48.5	48.5
k-NN (k=4)	50.0	52.7	56.7	50.0	50.0
k-NN (k=5)	51.6	52.3	56.2	51.6	51.6
Naive Bayes	57.7	57.0	58.7	57.3	57.3

5. Experimentos

Uma base de dados contendo 10 classes (blues, classical, country, disco, hiphop, jazz, metal, pop, reggae e rock) e 100 instâncias (ou músicas) de cada classe, totalizando 1.000 instâncias foi utilizada neste trabalho. Essa base foi a mesma utilizada nos experimentos de Tzanetakis e Cook [Tzanetakis and Cook, 2002][1]. Além dessa base padrão que contém as instâncias ordenadas por gênero, foi utilizada uma nova versão desta base com as instâncias de entrada ordenadas aleatoriamente no intuito de verificar a robustez dos métodos de meta-aprendizagem. Essa variação da base padrão foi chamada de base aleatória. É importante ressaltar que apesar da base ser utilizada ser a mesma que a utilizada em experimentos anteriores, uma comparação direta não é possível em virtude das diferenças nas condições experimentais.

Os experimentos foram realizados e avaliados com o método de validação-cruzada estratificada fator 10 (*Ten-Fold Stratified Cross-Validation*) e os resultados apresentados foram calculados utilizando a média das *F-measures* (calculada como a média harmônica dos valores de precisão e recobrimento) do resultado de cada classe. A precisão é o quociente entre o número de músicas que tiveram seu gênero classificado corretamente e o número de músicas que foram classificadas como sendo daquele gênero. O recobrimento é o quociente entre o número de músicas que tiveram seu gênero classificado corretamente e o número de músicas que possuem esse gênero.

No intuito de avaliar a eficácia de cada método de meta-aprendizagem, num primeiro momento, os experimentos foram realizados utilizando apenas um único classificador. Os resultados obtidos por cada classificador é apresentado na coluna *Individual* da tabela 1 para a base padrão e na tabela 2 para a base aleatória.

Os experimentos utilizando *Bagging* foram realizados visando determinar dois

[1]Disponível em: http://opihi.cs.uvic.ca/sound/genres

parâmetros: o tamanho da cesta (*bag size*) que é a porcentagem de exemplos que devem ser selecionados aleatoriamente para compor cada um dos n' conjuntos de treinamentos e o número de classificadores componentes que deveriam ser utilizados. Para o tamanho da cesta foram utilizados os valores de 50%, 60%, 70%, 80% e 90%; e para avaliar o número de classificadores foram utilizados: 5, 10, 20, 30, 40 e 50 classificadores componentes em cada experimento. As colunas *Pior Bag* e *Melhor Bag* das tabelas 1 e 2 apresentam o pior e o melhor resultado obtido utilizando *Bagging* para a base padrão e base aleatória respectivamente. Nos experimentos utilizando *Boosting* foram utilizados o mesmo número de classificadores componentes utilizados nos experimentos de Bagging, ou seja: 5, 10, 20, 30, 40 e 50. As colunas *Pior Boost* e *Melhor Boost* das tabelas 1 e 2 apresentam o pior e o melhor resultado obtidos utilizando *Boosting* para a base padrão e base aleatória respectivamente.

Avaliação Geral dos Resultados

Os resultados obtidos utilizando apenas um único classificador para a tarefa de classificação automática de gêneros musicais mostram que em ambas as bases o classificador Naive Bayes apresenta uma taxa de classificação superior aos demais. Por outro lado, o método de *Bagging* apresentou um bom desempenho para os classificadores J4.8 e k-NN e resultados não tão bons para o classificador Naive Bayes.

No caso do J4.8 o método apresentou um desempenho surpreendentemente positiva. No pior caso teve um aumento na taxa de acerto de 5% (utilizando um *bag size* de 50% e 5 classificadores) e no melhor caso o aumento foi de 14.8% (utilizando um *bag size* de 90% e 50 classificadores) para a base padrão; e aumento de 7.8% no pior caso (utilizando um *bag size* de 50% e 5 classificadores) e de 17.4% no melhor caso (utilizando um *bag size* de 80% e 30 ou 40 classificadores) para a base aleatória.

Já os resultados obtidos pelo método de *Bagging* aplicado ao classificador k-NN apresentaram uma melhora razoável em relação ao uso de um único classificador k-NN. Os resultados obtidos mostram que o desempenho mínimo obtido utilizando a técnica de *Bagging* aplicada ao classificador k-NN obteve resultados equivalentes ao uso de um único classificador k-NN. As únicas exceções foram para 1-NN na base padrão e 3-NN na base aleatória. porém apresentando uma melhora de até 5% no melhor caso. Um fato interessante é que na base padrão, com exceção do classificador 2-NN, o menor resultado foi obtido utilizando o mesmo *bag size* (70%) e em todos os casos os pior resultado foi obtido utilizando o mesmo número de classificadores (5), assim como em quase todos os casos os melhores resultados foram obtidos utilizando um *bag size* de 50%. Já na base aleatória os melhores resultados sempre foram obtidos utilizando um *bag size* de 50% com pelo menos 30 classificadores enquanto os piores resultados foram sempre obtidos utilizando 5 classificadores e com um *bag size* de 80% ou 90%. Dessa forma é possível perceber uma relação no caso do k-NN que parece indicar que os melhores resultados são normalmente obtidos com um *bag size* de 50%, enquanto que os piores resultados sempre são obtidos com um *bag size* de 70% ou superior.

Os resultados apresentados nas tabelas 1 e 2 mostram que o uso do método de *Bagging* aplicado ao classificador Naive Bayes se mostrou praticamente ineficiente. No primeiro caso, utilizando a base padrão, o resultado mínimo foi de 57.2% utilizando um *bag size* de 60% e 5 classificadores. Esse valor é 0.2% maior que o resultado obtido por um único classificador Naive Bayes, que é de 57%. Já no melhor caso, o resultado máximo atingido foi de 59.4% utilizando um *bag size* de 50% e 40 classificadores. No segundo caso, utilizando a base aleatória, o resultado mínimo foi de 57% utilizando um *bag size* de 60% e 10 classificadores. Esse valor é menor do que o obtido utilizando

um único classificador Naive Bayes na mesma base, que é de 57.7%. E mesmo no melhor caso o desempenho obtido foi de 58.7% utilizando um *bag size* de 50% e 30 ou 40 classificadores.

O método de *Boosting* utilizando o algoritmo AdaBoostM1 apresentou ganhos significativos para o classificador J4.8 e se mostrou ineficiente para os classificadores k-NN e Naive Bayes. Utilizando o J4.8 o método de *Boosting* apresentou uma performance surpreendente. No pior caso teve um aumento na taxa de acerto de 7.8% (utilizando cinco classificadores) e no melhor caso o aumento foi de 16% (utilizando 30 classificadores) para a base padrão; e aumento de 8.5% (utilizando cinco classificadores) no pior caso e de 15.8% (utilizando 50 classificadores) no melhor caso para a base aleatória.

Para o k-NN o método de *Boosting* se mostrou ineficiente, apresentando resultados aquém do esperando, normalmente piorando o desempenho do classificador ou em alguns raros casos aumentando ligeiramente o desempenho em no máximo 2%. Os resultados obtidos tanto na base padrão quanto na base aleatória mostram que o 2-NN foi o único caso em que o desempenho do pior caso foi superior a do classificador Individual, sendo o desempenho normalmente igual ou inferior ao uso de um único classificador. Já no melhor caso apenas o classificadores de 1-NN e 2-NN apresentam alguma melhora, enquanto nos demais casos, os melhores resultados obtidos são piores do que os resultados obtidos utilizando um único classificador k-NN. Dessa forma, o método de *Boosting* utilizando o classificador k-NN teve um desempenho aquém da esperada, não apresentando nenhuma melhora significativa no desempenho do classificador e em alguns casos até mesmo piorando o desempenho do mesmo quando comparada a um único classificador sem o uso da técnica de *Boosting*.

Para o Naive Bayes, os resultados obtidos com o método de *Boosting* foram aquém do esperado, piorando o desempenho do classificador em todos os casos.

6. Conclusões

Neste trabalho foi proposta uma nova abordagem para o problema da classificação automática de gêneros musicais utilizando as técnicas de meta-aprendizagem conhecidas como *Bagging* e *Boosting*. O método de *Bagging* se mostrou eficiente para os classificadores J4.8 e k-NN e ineficiente para o Naive Bayes. No caso do J4.8 a performance aumentou significativamente, enquanto que no caso do k-NN o uso da técnica de *Bagging* obteve normalmente nos piores resultados uma performance similar ao uso de um único classificador. No caso do Naive Bayes o método não apresentou melhora significativa e chegou inclusive a obter resultados inferiores ao uso de um único classificador.

O método de *Boosting* se mostrou eficiente para o classificador J4.8 e ineficiente para os classificadores k-NN e Naive Bayes. No caso do J4.8 a performance aumentou significativamente (16% no melhor caso), enquanto que para o k-NN o método normalmente piorou a performance do classificador tendo alguns raros casos onde a performance foi ligeiramente aumentada em no máximo 2%. Os resultados obtidos com o classificador Naive Bayes foram aquém do esperado piorando o desempenho do classificador em todos os casos.

Sintetizando os resultados obtidos, no caso de um único classificador o melhor desempenho foi a do classificador Naive Bayes que teve um acerto de 57% na base padrão e 57.7% na base aleatória. Já no caso das técnicas de meta-aprendizagem, os melhores resultados na base padrão foram obtidos com o uso de *Boosting* com um conjunto de classificadores J4.8 tendo um acerto no melhor caso de 68.1% e na base aleatória com o uso de *Bagging* tendo um acerto no melhor caso de 67.5%.

Como trabalho futuro os autores pretendem verificar o uso das técnicas de *Bagging* e *Boosting* aliadas a outros classificadores, além de rodar experimentos em outras bases contendo outros gêneros musicais.

Referências

Aha, D. W., Kidbler, D., and Albert, M. (1991). Instance-based learning algorithms. *Machine Learning*, 6(1):37–66.

Aucouturier, J. J. and Pachet, F. (2003). Representing musical genre: A state of the art. *Journal of New Music Research*, 32(1):83–93.

Breiman, L. (1996). Bagging predictors. *Machine Learning*, 24(2):123–140.

Carey, M. J., Parris, E. S., and Lloyd-Thomas, H. (1999). A comparison of features for speech, music discrimination. In *IEEE International Conference on Acoustics, Speech, and Signal Processing*, volume 1, pages 149–152.

Fingerhut, M. (1999). The ircam multimedia library: A digital music library. In *IEEE Forum on Research and Technology Advances in Digital Libraries*, pages 19–21.

Foote, J. T. A. (1999). An overview of audio information retrieval. *Multimedia Systems*, 7(1):42–51.

Freund, Y. and Schapire, R. E. (1996). Experiments with a new boosting algorithm. In *International Conference on Machine Learning*, pages 148–156.

Guo, G. and Li, S. Z. (2003). Content–based audio classification and retrieval by support vector machines. *IEEE Transactions on Neural Networks*, 14(1):209–215.

Kosina, K. (2002). Music genre recognition. Technical report, Fachlochschul Hagenberg.

Liu, T., Ogihara, M., and Li, Q. (2003). A comparative study on content-based music genre classification. In *Proc. of the 26th Annual Intl ACM SIGIR Conf. on Research and Development in Information Retrieval*, pages 282–289, Toronto, Canada.

Mitchell, T. M. (1997). *Machine Learning*. McGraw-Hill.

Pampalk, E., Rauber, A., and Merkl, D. (2002). Content–based organization and visualization of music archives. In *ACM Multimedia 2002*, pages 570–579, Juan-les-Pins, France.

Pye, D. (2000). Content–based methods for the management of digital music. In *IEEE International Conference on Acoustics, Speech and Signal Processing*, pages 2437–2440.

Quinlan, J. R. (1993). *C4.5: Programs for Machine Learning*. Morgan Kaufmann.

Shao, X., Xu, C., and Kankanhalli, M. S. (2003). Applying neural network on the content–based audio classification. In *Fourth International Conference on Information, Communications and Signal Processing*, volume 3, pages 1821–1825.

Sweldens, W. and Piessens, R. (1993). Wavelet sampling techniques. In *Proc. of the Statistical Computing Section*, pages 20–29.

Tzanetakis, G. and Cook, P. (2002). Musical genre classification of audio signals. *IEEE Transactions on Speech and Audio Processing*, 10(5):293–302.

Zhang, T. and Kuo, C. C. J. (2001). Audio content analysis for online audiovisual data segmentation and classification. *IEEE Transactions on Speech and Audio Processing*, 9(4):441–457.