

Classificação Automática de Gêneros Musicais

Utilizando Métodos de Bagging e Boosting

Carlos N. Silla Jr. Celso Kaestner Alessandro Koerich

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

10º Simpósio Brasileiro de Computação Musical
Belo Horizonte - Minas Gerais - Brasil
3 a 6 de Outubro, 2005

Quantidade de Conteúdo Multimídia Disponível On-line

- ▶ Necessidade de ferramentas capazes de organizar e gerenciar essa grande quantidade de informações;
- ▶ A música digital é 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.

Organização de Arquivos Multimídia

Estratégia Usual

- ▶ Baseada em Meta-Informações Textuais;
- ▶ Ex: Músicas no formato MP3 utilizam rótulos ID3.

Limitação

- ▶ São geradas manualmente para depois serem associadas ao arquivo multimídia.

Solução

- ▶ Preencher essas informações através da análise do conteúdo do arquivo multimídia.

Gêneros Musicais

- ▶ Descrição importante que tem sido utilizada para classificar e caracterizar músicas digitais e para organizar grandes coleções disponíveis na Internet;
- ▶ São rótulos categóricos criados por especialistas humanos assim como por amadores;
- ▶ Relacionados com a instrumentalização utilizada, estrutura rítmica e conteúdo harmônico da música.

Problema

- ▶ Conceito relativamente subjetivo, e mesmo a indústria musical muitas vezes é contraditória ao atribuir gêneros musicais para as músicas;

Gêneros Musicais

- ▶ Uma prática muito comum 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;
- ▶ 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.

Visão Geral do Sistema



Extração de Características

São extraídas três tipos de características diretamente dos primeiros 30 segundos do sinal da música [Tzanetakis and Cook, 2002]:

- ▶ Textura Timbral (*Timbral texture*);
- ▶ Relacionadas à batida (*beat-related*);
- ▶ Relacionadas as variações da frequência da vibração (*pitch-related*);

Textura Timbral

Calculadas com base na Transformada de Fourier. São extraídas 19 características:

- ▶ Média e Variância do Centróide Espectral (*Spectral Centroid*);
- ▶ Média e Variância do Rolloff Espectral (*Spectral Rolloff*);
- ▶ Média e Variância do Fluxo Espectral (*Spectral Flux*);
- ▶ Média e Variância das Taxas de Cruzamento Zero (*Time Domain Zero-Crossings*);
- ▶ Médias e Variâncias dos Cinco Primeiros Coeficientes cepstrais de frequência-Mel (MFCC);
- ▶ Baixa Energia (*Low Energy*).

Relacionadas à batida

Calculadas com base em um histograma de batidas. São extraídas 6 características:

- ▶ Amplitude Relativa do Primeiro Pico;
- ▶ Amplitude Relativa do Segundo Pico;
- ▶ A razão da amplitude do segundo pico dividida pela amplitude do primeiro pico;
- ▶ O período do primeiro pico em BPM;
- ▶ O período do segundo pico em BPM;
- ▶ A soma do histograma;

Relacionadas ao Pitch

Calculadas com base em duas versões de um histograma de Pitch.
São extraídas 5 características:

- ▶ Amplitude do maior pico do primeiro histograma.
- ▶ Período do pico máximo do primeiro histograma.
- ▶ Período do pico máximo do segundo histograma.
- ▶ O intervalo Pitch entre os dois picos mais proeminentes do primeiro histograma.
- ▶ A soma do histograma.

Classificação Utilizando Métodos de Meta-Aprendizagem

Métodos de Meta-Aprendizagem

- ▶ *Bagging* [Breiman, 1996]
- ▶ *Boosting* [Mitchell, 1997]

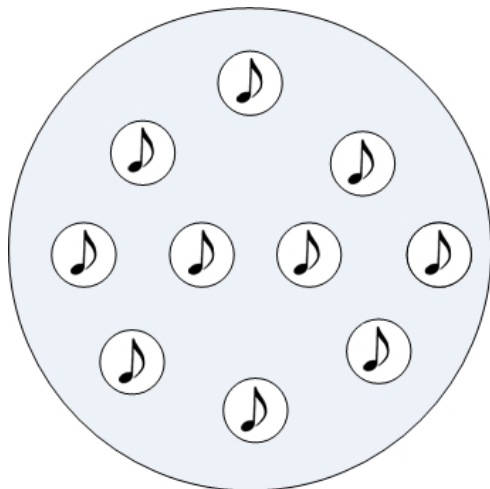
Classificadores

- ▶ Árvores de Decisão (J4.8) [Quinlan, 1993]
- ▶ k-NN (k vizinhos mais próximos) [Aha et al., 1991]
- ▶ Naive Bayes [Mitchell, 1997]

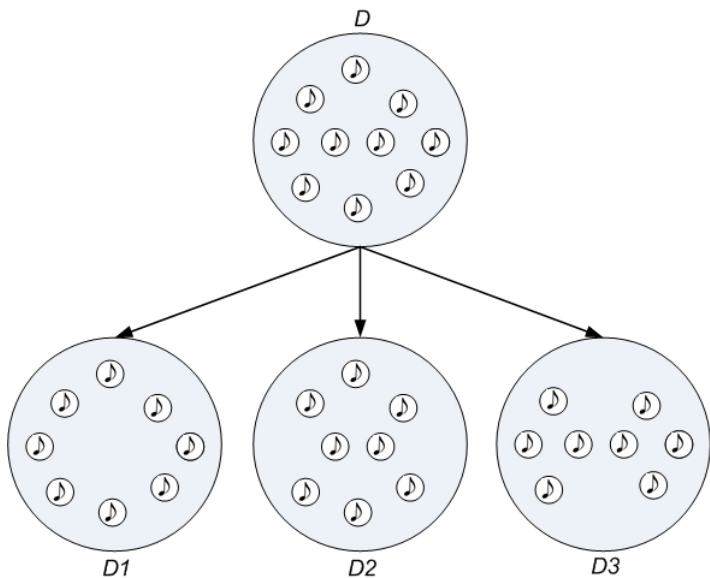
Funcionamento do Algoritmo de *Bagging*

- ▶ 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”.
- ▶ A decisão da classificação final é baseada no voto de cada componente.
- ▶ Dessa forma, recebe dois parâmetros:
 1. Tamanho do Cesto;
 2. Número de Classificadores;

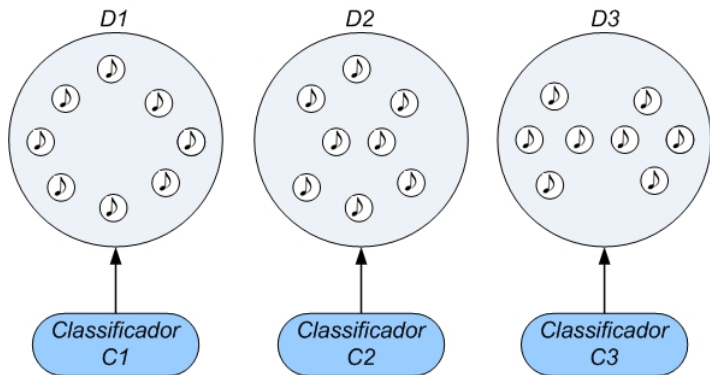
Exemplo: Tamanho do Cesto = 80% Classificadores = 3



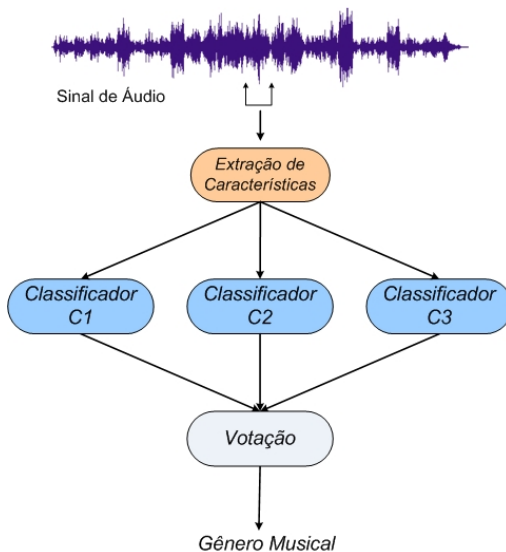
Conjunto de Treinamento D



Treinamento dos Classificadores Componentes



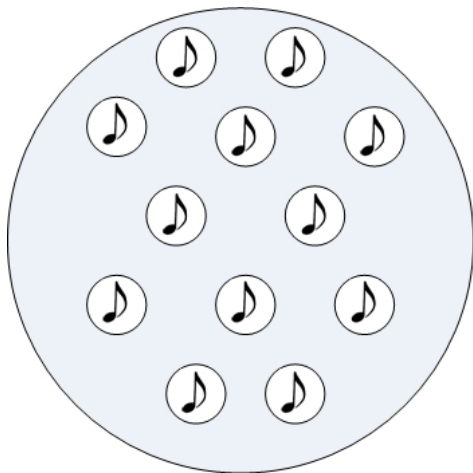
Teste de uma nova música



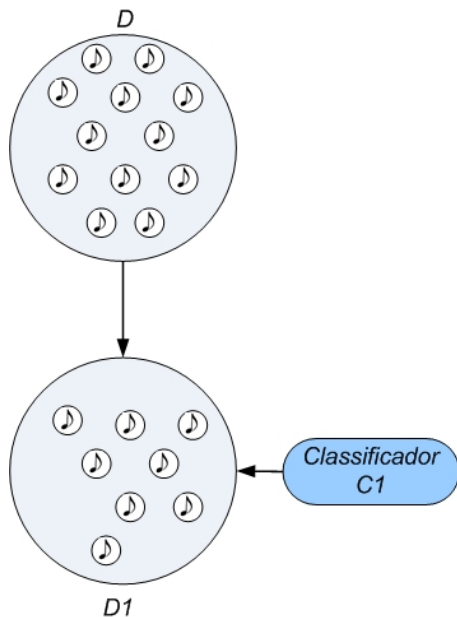
Funcionamento do Algoritmo de *Boosting*

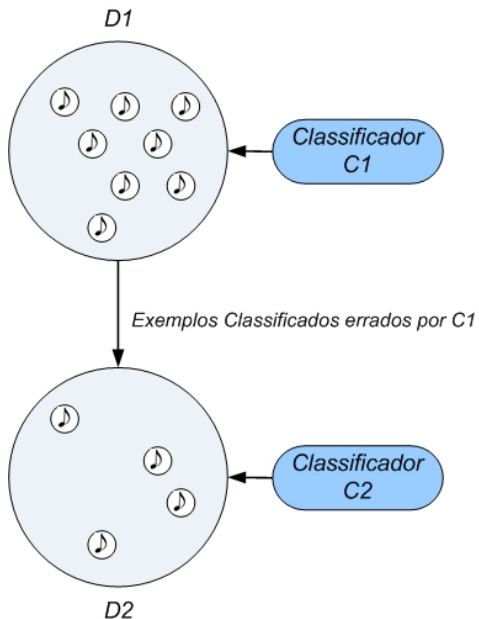
- ▶ Tem como objetivo melhorar a precisão de qualquer algoritmo de aprendizagem.
- ▶ O procedimento utilizado é:
 1. Seleciona-se aleatoriamente $n' < n$ amostras do conjunto de treinamento D , sem reposição.
 2. É criado um classificador com precisão sobre o conjunto de treinamento maior do que a média;
 3. Adicionalmente novos “classificadores componentes” são adicionados para formar um conjunto cuja regra de decisão tenha uma alta precisão arbitrária sobre o conjunto de treinamento.
- ▶ Recebe um único parâmetro: Número de Classificadores;

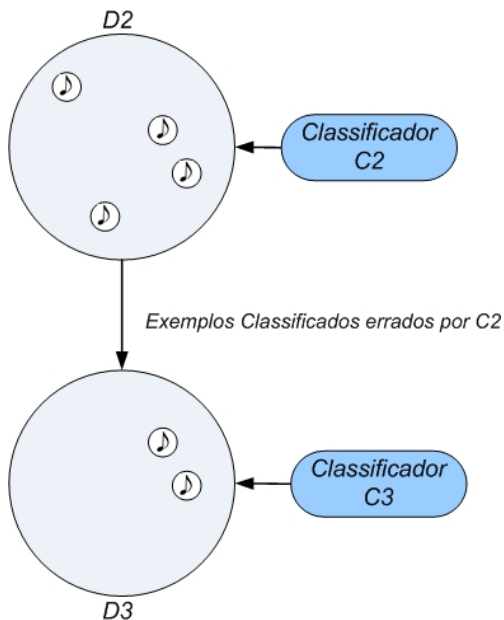
Exemplo: Classificadores = 3



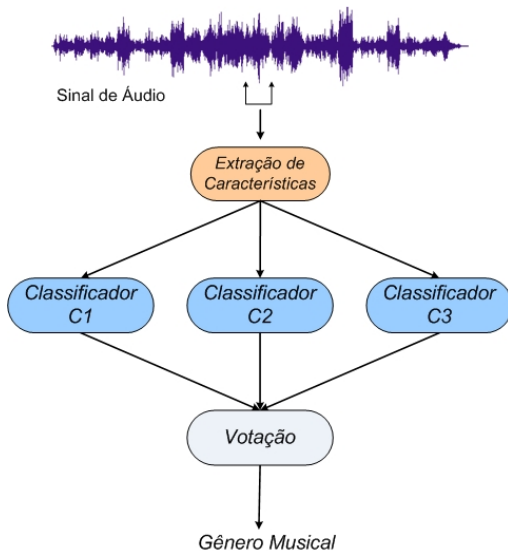
Conjunto de Treinamento D







Teste de uma nova música



Base de Dados

Dados Utilizados

- ▶ 1.000 Músicas (Instâncias) de 10 Gêneros (Classes): Blues, Classical, Country, Disco, Hiphop, Jazz, Metal, Pop, Reggae e Rock.
- ▶ Mesma base utilizada nos experimentos de [Tzanetakis and Cook, 2002];
- ▶ Disponível em: <http://opihi.cs.uvic.ca/sound/genres>

Bases Geradas

- ▶ Base Padrão: Contém as instâncias ordenadas por gênero;
- ▶ Base Aleatória: Contém as instâncias ordenadas aleatoriamente;

Método de Avaliação

- ▶ Validação Cruzada Estratificada Fator 10 (*Ten-Fold Stratified Cross-Validation*).
- ▶ Os resultados foram calculados utilizando a média das *F-measures* (média harmônica dos valores de precisão e recobrimento) do resultado de cada classe.

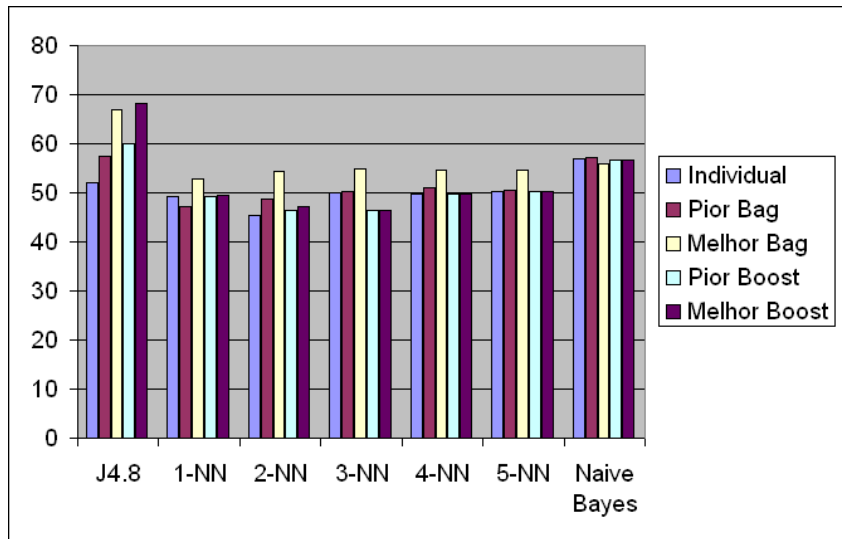
Parâmetros Utilizados para *Bagging*

- ▶ Tamanho da Cesta (*Bag Size*): 50%, 60%, 70%, 80%, 90%;
- ▶ Número de Classificadores: 5, 10, 20, 30, 40, 50;

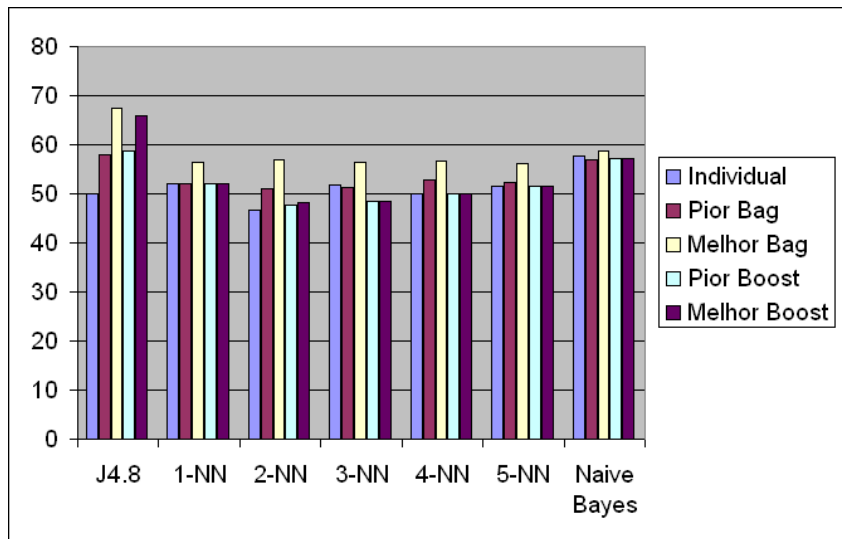
Parâmetros Utilizados para *Boosting*

- ▶ Número de Classificadores: 5, 10, 20, 30, 40, 50;

Resultados – Base Padrão



Resultados – Base Aleatória



Bagging

Avaliação

- ▶ J4.8: A performance aumentou significativamente.
- ▶ k-NN: Piores resultados performance similar a um único classificador.
- ▶ Naive Bayes: Não apresentou melhora significativa. Obteve inclusive resultados inferiores ao uso de um único classificador.

Boosting

Avaliação

- ▶ J4.8: A performance aumentou significativamente (16% no melhor caso).
- ▶ k-NN: Normalmente piorou a performance do classificador. Raros casos aumentou a performance em 2%.
- ▶ Naive Bayes: Aquém do esperado. Piorou a performance do classificador em todos os casos.

Síntese dos Resultados

Melhor performance

- ▶ Classificador Individual: Naive Bayes
 - ▶ Base Padrão: 57%
 - ▶ Base Aleatória: 57.7%
- ▶ Meta-Aprendizagem:
 - ▶ Base Padrão: 68.1% (*Boosting* – J4.8)
 - ▶ Base Aleatória: 67.5% (*Bagging* – J4.8)

Avaliação dos Métodos de Meta-Aprendizagem

Bagging

- ▶ Eficiente para: J4.8 e k-NN
- ▶ Ineficiente para: Naive Bayes






Boosting

- ▶ Eficiente para: J4.8
- ▶ Ineficiente para: k-NN e Naive Bayes

Trabalhos Futuros

- ▶ Uso das técnicas de *Bagging* e *Boosting* aliadas a outros classificadores.
- ▶ Experimentos em bases contendo outros gêneros musicais.
- ▶ Utilizar mais de um segmento da mesma música;

Referências

-  Aha, D. W., Kidbler, D., and Albert, M. (1991). Instance-based learning algorithms. *Machine Learning*, 6(1):37–66.
-  Breiman, L. (1996). Bagging predictors. *Machine Learning*, 24(2):123–140.
-  Mitchell, T. M. (1997). *Machine Learning*. McGraw-Hill.
-  Quinlan, J. R. (1993). *C4.5: Programs for Machine Learning*. Morgan Kaufmann.
-  Tzanetakis, G. and Cook, P. (2002). Musical genre classification of audio signals. *IEEE Transactions on Speech and Audio Processing*, 10(5):293–302.

Perguntas?

