SISTEMAS DE CLASSIFICAÇÃO MUSICAL COM REDES NEURONAIS

Ricardo Malheiro*/*****
Rui Pedro Paiva**
António José Mendes**
Teresa Mendes**
Amílcar Cardoso**

Como resultado da evolução e inovação tecnológicas, a indústria da distribuição electrónica de música tem tido um enorme crescimento. Desta forma, tarefas como a classificação automática de géneros musicais tornam-se um fortíssimo motivo para o incremento da investigação na área. O reconhecimento automático de géneros musicais envolve tarefas como a extracção de características das músicas e o desenvolvimento de classificadores que utilizem essas características. Neste estudo pretendeu-se, através de 3 problemas de classificação independentes, classificar segmentos de música clássica em subgéneros. Para tal, foram extraídas 40 características por segmento musical, tendo os classificadores utilizados sido redes neuronais. Devido à qualidade dos resultados alcançados, foi construído em seguida um protótipo para um sistema real de classificação. Neste, de um conjunto de músicas não catalogadas, foram extraídos dez segmentos às "cegas" que foram classificados utilizando os classificadores anteriores. Cada música foi classificada no género mais representado pelos seus segmentos.

Palavras Chave: redes neuronais, music information retrieval, classificação de música, extracção de características, análise de sinal musical

1. Introdução

A classificação de música tem-se tornado para o Homem cada vez mais importante, à medida que aumentam a quantidade de música disponível e as necessidades de catalogação. Essa catalogação pode tomar as mais diversas formas: por género musical, por artista, por época, por nacionalidade, por instrumento(s), por tipo de voz (e.g., feminina), por contexto em que se insere (e.g., música pertencente à banda sonora de um filme), etc. Desde sempre a classificação tem sido feita manualmente, o que acarreta alguns problemas óbvios, como por exemplo o tempo que é necessário para concluir essa operação caso os dados a organizar sejam volumosos.

* Universidade Católica Portuguesa – Centro Regional das Beiras – Viseu
** Centro de Informática e Sistemas da Universidade de Coimbra (CISUC)

Outro problema é o grau de subjectividade associado a cada classificação, já que os resultados desta dependem da pessoa que classifica, do seu conhecimento musical e da sua experiência.

Nos últimos anos, devido a vários factores como por exemplo, a evolução rápida dos computadores tanto a nível de software como hardware, o aumento gradual e generalizado de largura de banda disponível, a universalização cada vez maior da Internet e a criação de novos formatos de compressão de música mais eficientes (e.g., MP3), a Internet tornou-se rapidamente um mercado apetecível para um determinado conjunto de serviços que tiram partido desses factores para aumentar a satisfação e a comodidade dos utilizadores.

Estes serviços, como por exemplo, os de compra e venda de música (e.g., sítios como *AllMusicGuide* – www.allmusic.com, ou *CDNOW* – www.cdnow.com) necessitam, para se tornarem interessantes para os utilizadores, de bases de dados de música sempre actualizadas e motores de pesquisa em tempo real eficientes e rápidos que respondam sempre que possível com sucesso às pesquisas desses utilizadores.

Para se conseguir o sucesso nas pesquisas a essas bases de dados, é necessário organizá-las segundo taxonomias que vão de encontro às necessidades dos utilizadores. Uma das taxonomias mais comuns na classificação de música consiste na hierarquização de géneros musicais. Assim sendo, cada nova música a inserir na base de dados, deverá ser previamente classificada num dos géneros da taxonomia em causa, tarefa esta por vezes bastante subjectiva. Isto fará, como é óbvio, com que o processo de actualização das bases de dados se torne ainda mais moroso e complexo. Esta consequência, juntamente com o facto de todos os dias serem adicionadas milhares de novas músicas nas bases de dados, faz com que os métodos manuais de classificação sejam ineficazes na resposta a essas necessidades. Assim, surge a inevitabilidade da utilização do próprio computador para esse tipo de tarefas, através de sistemas de classificação automática.

Este tipo de problemas tem sido estudado recentemente por vários investigadores.

George Tzanetakis e Perry Cook em [Tzanetakis & Cook, 2002] classificam música em 10 géneros musicais, nomeadamente música clássica, *country*, disco, *hip-hop*, *jazz*, *rock*, *blues*, *reggae*, *pop* e metal. Especializam ainda os classificadores utilizados em dois tipos de música: música *jazz* e música clássica. No caso do *jazz*, consideram 6 subgéneros; *bigband*, *cool*, fusão, piano, quartetos e *swing* e no caso da música clássica

4 subgéneros; música coral, orquestra, piano e quarteto de cordas. As características tímbricas utilizadas são: centróide, *rolloff*[1], *flux*[2], MFCC (*Mel-Frequency Cepstral Coefficients*) e *zcr*[3]. Estas características são calculadas em janelas de análise de pequena duração (23ms). Posteriormente são calculadas as médias e variâncias das características anteriores em intervalos de tempo de 1s. É calculada ainda a característica *low-energy* nos mesmos intervalos de 1s. As características rítmicas usadas são calculadas a partir do histograma de batidas da música. As características do conteúdo de *pitch*[4] são baseadas em técnicas de detecção de múltiplos *pitches*. Tanto as características rítmicas como as de conteúdo de *pitch* são calculadas em relação a toda a música. No sentido de avaliar a importância das características foram utilizados dois classificadores: Modelos de Misturas Gaussianas (GMM)[5] e *K*-Vizinhos mais Próximos (KNN)[6]. Os resultados de classificação alcançados foram de 61% para os 10 géneros e 82,25% para a música clássica. Outro estudo com algumas variantes, dos mesmos autores juntamente com Georg Essl, foi publicado em [Tzanetakis et al., 2001].

Seth Golub [Golub, 2000] classifica música em sete géneros bastante diferentes: *a cappela*, celta, clássica, electrónica, jazz, latina e *pop-rock*. As características utilizadas são *loudness*[7], centróide, largura de banda e uniformidade, bem como outras características estatísticas obtidas a partir delas. Foram utilizados três classificadores: Modelo linear generalizado (GLM)[8], MLP[9] e KNN. Os melhores resultados de classificação conseguidos foram de 67%.

Karin Kosina [Kosina, 2002] classifica em apenas três géneros muito diferentes: metal, dança e clássica. Utiliza um conjunto de características que engloba MFCC, *zcr*, energia e batida. Foi conseguida uma taxa de sucesso na classificação de 88% para o classificador utilizado, KNN.

Keith Martin [Martin, 1998] e [Martin & Kim, 1998] estuda igualmente o problema da identificação de instrumentos, propondo um conjunto de características relacionadas com as propriedades físicas dos instrumentos com o objectivo de os identificar num

[1] Medida da forma do espectro do sinal
[2] Medida da alteração do espectro do sinal
[3] Medida do conteúdo de frequência do sinal
[4] Percepção que o ouvido humano tem da frequência do sinal
[5] Em terminologia Inglesa: Gaussian Mixture Models
[6] Em terminologia Inglesa: *K*-Nearest Neighbors
[7] Percepção que o ouvido humano tem da intensidade do som
[8] Em terminologia Inglesa: Generalized Linear Model
[9] Em terminologia Inglesa: Multilayer Perceptron

ambiente polifónico. Outros estudos ainda sobre identificação de instrumentos foram publicados em [Fraser & Fujinaga, 1999].

O objectivo do nosso estudo é a classificação de música em subgéneros da música clássica, nomeadamente ópera, música coral e música para flauta, piano e violino. Esta escolha deve-se ao facto de não existirem muitos estudos especificamente sobre música clássica. Além do mais as bases de dados de música digital têm uma grande diversidade de taxonomias de música clássica, o que demonstra a utilidade prática do estudo.

Ao contrário de alguns investigadores que optaram por géneros bastante díspares (e.g., disco, música clássica, *jazz*, *rock*, etc), nós escolhemos um conjunto de géneros musicais bastante similares divididos em vários problemas de classificação: discriminação de música instrumental, de música vocal e de música clássica no seu todo. Esta similaridade entre os géneros musicais deverá conduzir a uma maior complexidade no que concerne aos três problemas de classificação em estudo.

Foram escolhidas características baseadas nas utilizadas em [Tzanetakis et al., 2001] e [Golub, 2000]: *zcr*, *loudness*, centróide, largura de banda e uniformidade. Estas características privilegiam a análise do conteúdo de *pitch* e de timbre do sinal, o que parece ser o mais apropriado tendo em conta os problemas de classificação em análise.

Foi utilizado como classificador, para cada problema, uma rede MLP treinada com o algoritmo de Levenberg-Marquardt. Os resultados alcançados pelos classificadores anteriores serviram ainda de base para a construção de sistemas de classificação automática.

Este artigo está organizado da seguinte forma. A Secção 2 descreve o processo de extracção de características e as características utilizadas. Na Secção 3 é apresentada uma breve introdução teórica de redes neuronais, mais especificamente de redes neuronais com ligações para a frente. É ainda descrita a sua aplicação aos nossos problemas de classificação em géneros musicais. Os resultados experimentais são apresentados e analisados na Secção 4. Na Secção 5 é definido um protótipo de um sistema automático de reconhecimento de géneros musicais. São também analisados os seus resultados. Finalmente, na Secção 6 são extraídas algumas conclusões sobre o trabalho realizado, bem como direcções possíveis para trabalho futuro.

2. Extracção de Características

De acordo com os tipos de classificação a efectuar, baseados essencialmente na discriminação entre instrumentos e na discriminação entre voz e parte instrumental e tendo como base trabalhos efectuados por outros autores como Golub [Golub, 2000] e Tzanetakis [Tzanetakis & Cook, 2002] as características extraídas de cada peça musical foram escolhidas de forma a privilegiar a análise do timbre e do *pitch* do sinal. Não foram utilizadas características rítmicas já que não pareceram relevantes para o problema de classificação em questão. No entanto, a intenção é utilizá-las futuramente para as avaliar neste contexto.

Começou-se por seleccionar um extracto musical de 6s (frequência de amostragem de 22 kHz, quantização de 16 bits, monoaural) por cada uma das peças musicais utilizadas. No processo de treino de uma rede neuronal, o conjunto de amostras utilizadas deve ter o mínimo de ambiguidade possível em relação à classe a que pertencem. Por esse facto foram escolhidos extractos musicais considerados relevantes para cada um dos géneros musicais em análise. O objectivo deste estudo não é construir um sistema que utilize trechos de música de longa duração, mas sim de pequena duração e significativos para cada um dos géneros musicais, pelo que foram seleccionados extractos de curta duração. A ideia é imitar de alguma maneira, a forma como os seres humanos classificam a música [Perrot & Gjerdigen, 1999], i.e. conseguir classificar utilizando pequenos segmentos de música e usando apenas características extraídas directamente da análise de superfície feita ao sinal.

O processo de extracção de características começa, para cada uma das músicas representadas, por dividir o sinal de 6s em janelas de 23,22 ms com 50% de sobreposição entre duas janelas consecutivas. Este comprimento específico de janela foi escolhido por forma ao número de amostras em cada janela ser potência de 2, o que é fundamental para optimizar a eficiência da Transformada Rápida de Fourier (FFT)[10] [Smith, 1997]. Isto dá-nos um total de 512 amostras por janela num total de 515 janelas. Para cada janela, o sinal é então multiplicado pela função de Hanning, que é caracterizada por um bom compromisso entre a resolução espectral e o esbatimento espectral [Smith, 1997].

[10] Em terminologia Inglesa: Fast Fourier Transform

Neste ponto, são extraídas directamente do sinal, para cada janela, o *loudness* e o *zcr*. São portanto extraídas no domínio do tempo.

Após a aplicação da FFT a cada janela, são extraídas três características espectrais: o centróide, a largura de banda e a uniformidade. A partir destas cinco características base são calculadas por processos estatísticos todas as 40 características que irão representar cada peça de música.

Nos 6 s de cada peça musical, existem portanto 132300 (6x22050) amostras. Sabendo que cada janela tem 23,22 ms e estas têm 50% de sobreposição, existem no total 512 amostras (512 x período de amostragem do sinal = 0,02322 s) por janela de um total de 515 janelas ((132300 - 512/2) / (512/2) = 515). Em cada janela, a resolução a nível de frequência[11] é de 43,06 Hz, i.e., no domínio da frequência de uma amostra para a seguinte, existe um salto de 43,06 Hz.

Vai ser descrito em seguida todo o processo para o cálculo das características temporais e espectrais.

Características temporais

Foram utilizadas duas características extraídas no domínio do tempo. O *loudness* e o número de intersecções com o eixo das abcissas ou seja do tempo (*zcr*).

Loudness

É uma característica perceptual que tenta captar a percepção que o ouvido humano tem da intensidade do som. A informação extraída do sinal sonoro que vai servir de base ao cálculo do *loudness* é a amplitude.

Assim, o *loudness* pode ser representado pela equação seguinte (1):

$$L(r) = log_2\left(1 + \frac{1}{N}\sum_{n=1}^{N}|x(n)|\right) \quad (1)$$

[11] Resolução de frequência = (frequência de amostragem) / (número de amostras) [Smith, 1997].

onde *L* representa o *loudness*, *r* o número da janela actual, *N* é o número de amostras em cada janela, *n* é o número da amostra actual na janela actual e finalmente *x(n)* representa a amplitude da *n*-ésima amostra na janela actual.

ZCR

Esta característica mede simplesmente o número de vezes que o sinal sonoro atravessa o eixo das abcissas (tempo), sendo representada por (2):

$$Z(r) = \frac{1}{2}\sum_{n=1}^{N} \left| sgn(x(n)) - sgn(x(n-1)) \right| \quad (2)$$

Na expressão anterior, *Z(r)* representa o número de intersecções com o eixo das abcissas que existem na janela *r* e *sgn(x(n))* representa o sinal da amplitude da *n*-ésima amostra da janela *r*.

Esta é uma medida do conteúdo de frequência do sinal. É muitas vezes usada em problemas de discriminação entre música e voz e para determinar a quantidade de ruído de um sinal [Tzanetakis & Cook, 2002].

Características espectrais

As características espectrais utilizadas, calculadas no domínio da frequência, são o centróide, a largura de banda e a uniformidade.

O processo que culmina no cálculo das características espectrais anteriores começa pela conversão do sinal original para o domínio da frequência, utilizando para tal a Transformada de Fourier para pequenos segmentos (STFT)[12] [Polikar, 2003]. Desta forma o sinal original é dividido em janelas, como foi visto atrás. Para cada janela, o sinal é então multiplicado pela função de Hanning. Finalmente é calculada a FFT de cada janela.

[12] Em terminologia Inglesa: Short-Time Fourier Transform

Centróide

Esta característica espectral pode ser definida como a média pesada das magnitudes das frequências. É também um indicador do "brilho" do sinal [Wold et al., 1996]. Assim, valores altos para esta característica indicam um sinal com maior brilho e frequências globalmente mais altas para esse sinal. Wold explica claramente este conceito de brilho com uma experiência: se ao emitir um som, se puser a mão à frente da boca, vai-se diminuir o brilho, assim como o *loudness* do som.

Normalmente o centróide reflecte-se na voz por valores mais baixos e na música por valores mais altos, portanto é considerada uma característica fundamental para a discriminação entre voz e música.

Esta característica pode ser representada pela seguinte equação (3):

$$C(r) = \frac{1}{N} \frac{\sum_{k=1}^{N} M_r(k) . log_2 k}{\sum_{k=1}^{N} M_r(k)} \tag{3}$$

onde *C(r)* representa o valor do centróide na janela *r* e $M_r(k)$ representa a magnitude da transformada de Fourier na janela *r* e no índice de frequências *k*.

Largura de Banda

A definição desta característica espectral pode ser dada como a média pesada dos desvios padrões das bandas de frequência [Golub, 2000], ou muito simplesmente como desvio padrão da frequência. Se esta característica tiver um valor baixo, isso significa que as frequências do sinal concentram-se todas perto do centróide, i.e., há uma gama mais estreita de frequências no sinal.

Para se perceber melhor, uma sinusóide seno tem largura de banda igual a zero, enquanto um ruído tem normalmente uma largura de banda elevada.

A equação (4) permite calcular esta característica:

$$B(r) = \sqrt{\frac{\sum_{k=1}^{N} (C(r) - log_2\, k)^2 M_r(k)}{\sum_{k=1}^{N} M_r(k)}} \quad (4)$$

onde *B(r)* representa a largura de banda da janela *r*, *C(r)* é como verificamos atrás o centróide dessa mesma janela.

Uniformidade

A última característica espectral calculada foi a uniformidade, a qual mede a similaridade entre as magnitudes das bandas de frequência presentes no sinal.

Esta característica é fundamental para discriminar entre sinais com magnitudes muito altas para um reduzido número de frequências e sinais em que os valores das magnitudes sejam semelhantes para a grande maioria das frequências. No caso extremo, uma sinusóide tem uniformidade igual a zero, enquanto um sinal em que o ruído existente seja claro, deverá ter um valor elevado para esta característica.

Esta característica pode ser representada pela equação que se segue (5):

$$U(r) = -\sum_{k=1}^{N} \frac{M_r(k)}{\sum_{k=1}^{N} M_r(k)} . log_N \frac{M_r(k)}{\sum_{k=1}^{N} M_r(k)} \quad (5)$$

onde *U(r)* representa a uniformidade da janela *r*.

Primeiras Diferenças

Para cada janela são extraídas, as cinco características anteriores. Seguidamente, são calculadas as diferenças dos valores dessas cinco características entre janelas consecutivas, e.g., *L(r) - L(r-1)* para o caso do *loudness* em que *r* representa a janela corrente.

Estas cinco novas características juntamente com as cinco calculadas anteriormente constituem as 10 características base.

Assinatura de Cada Peça Musical

A música clássica, objecto deste estudo, caracteriza-se em geral por variações acentuadas ao longo do tempo, das características base descritas anteriormente. Por isso, pensa-se que as manipulações estatísticas em relação a essas características poderão influir na obtenção de bons resultados.

Para cada uma das 10 características base, são calculadas, de 2 em 2 segundos, as características intermédias. São constituídas pelas médias e pelos desvios padrões dos valores de cada característica base em todas as janelas em cada intervalo de 2 segundos. Como os extractos de música neste trabalho têm 6 segundos, isso quer dizer que estas características são calculadas 3 vezes para cada uma das características base. Portanto existem 20 (2x10) características intermédias para cada intervalo de 2s de sinal.

Calculando as médias e desvios padrões das características intermédias, chegamos às características que constituem no seu conjunto a representação de cada extracto musical. Essa representação é também chamada de assinatura do extracto musical. A assinatura é, portanto, constituída por 40 características (2x2x10).

3. Classificação com Redes Neuronais Artificiais

As redes neuronais artificiais (ANN)[13] são modelos computacionais criados com o intuito de emular o funcionamento do cérebro humano. Pretende-se, à imagem do cérebro, que as ANN tenham capacidade de aprendizagem, de adaptação e de generalização. Não obstante o ainda pouco conhecimento sobre determinados mecanismos que o cérebro utiliza e as limitações dos próprios computadores, sabe-se que para um largo espectro de problemas, as ANN conseguem aproximar uma solução com resultados muito satisfatórios.

São aplicadas em áreas tão diversas como reconhecimento de voz, robótica, investigação médica, telecomunicações, marketing, análise de investimentos, reconhecimento automático de géneros musicais, jogos, etc.

As redes neuronais com ligações para a frente (FFNN) constituem uma classe especial de ANN, nas quais todos os neurónios de uma determinada camada *l* estão ligados a todos os neurónios da camada *l-1* (Figura 1).

[13] Em terminologia Inglesa: Artificial Neural Networks

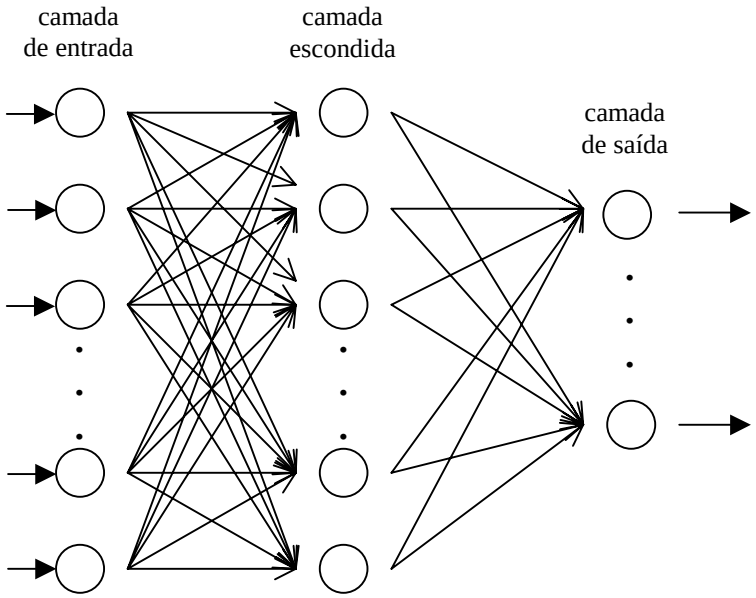


Figura 1. Rede neuronal com ligações para a frente

Como se verifica na figura, uma rede com ligações para a frente é constituída tipicamente por uma camada de entrada, que corresponde aos dados que entram na rede, uma camada escondida (podem ser utilizadas mais camadas [Sarle, 2001]) cujos neurónios recebem os dados produzidos pelos neurónios da camada de entrada e uma camada de saída, cujos neurónios recebem dados da camada escondida e que correspondem à saída da rede.

Como exemplos de redes deste tipo, encontramos as redes MLP (Perceptrão Multicamada) [Rumelhart & McClelland, 1986], as redes RBF (redes com função de base radial) [Broomhead & Lowe, 1988] e as redes LVQ[14] [Kohonen, 1989].

Neste estudo foram utilizadas especificamente as redes MLP, assim, todas as considerações futuras, serão referentes a este tipo de redes, apesar de algumas delas poderem ser válidas para outras redes com ligações para a frente.

A estrutura fundamental numa rede neuronal é o neurónio. Cada neurónio é estimulado ou seja, recebe sinais dos neurónios vizinhos, enviando sinais após processamento, para outros neurónios. Este processo de comunicação entre os neurónios assemelha-se muito ao que se passa com os neurónios biológicos e as sinapses.

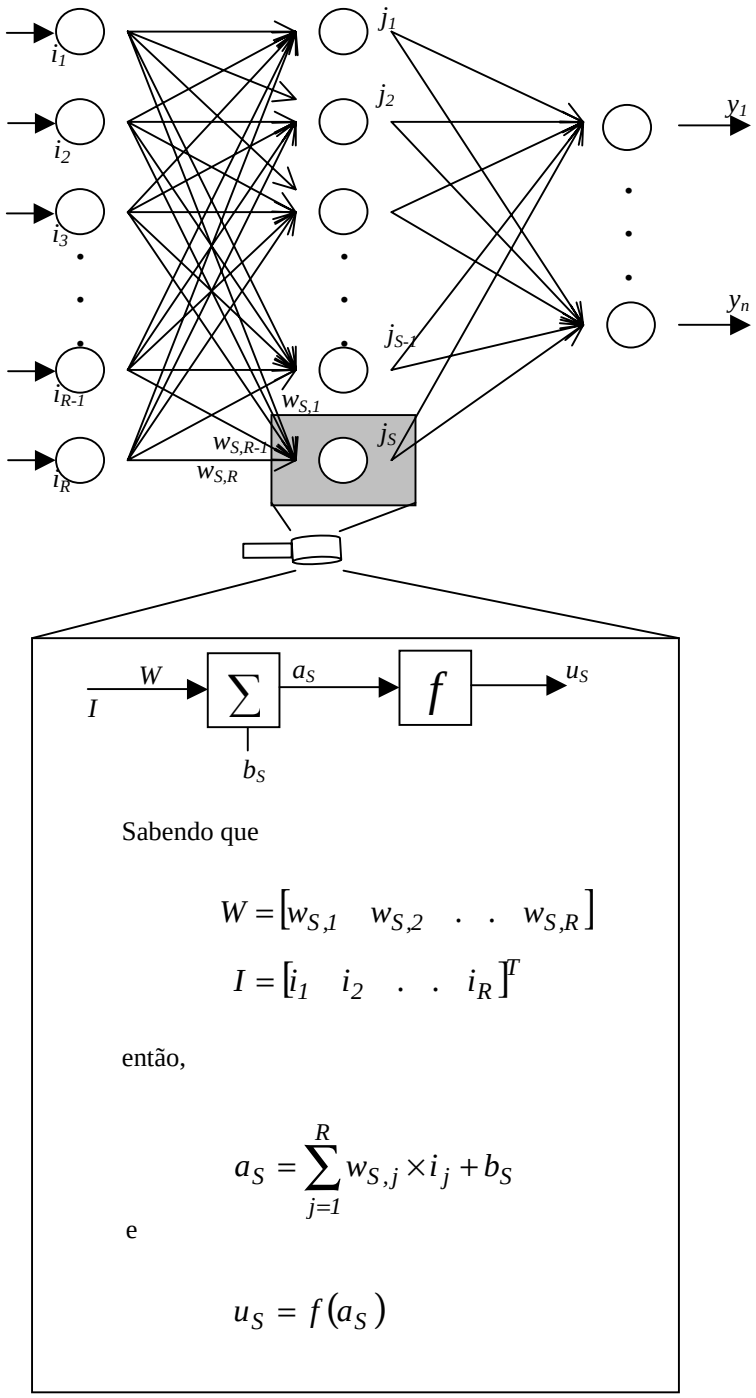
Pode-se visualizar na Figura 2, para um dado neurónio de uma rede neuronal, o processo de produção de um sinal a partir de sinais enviados pelos neurónios vizinhos e dos pesos que influenciam esse neurónio.

[14] em terminologia inglesa: Learning Vector Quantization

Os resultados de saída da rede neuronal dependem dos dados de entrada, dos valores iniciais dos parâmetros da rede e da relação entre os próprios neurónios. Essa relação, como se visualiza por exemplo na figura anterior para o *s*-ésimo neurónio da camada escondida, é representada pelo produto da matriz de pesos que incide nesse neurónio, *W* (e.g., o elemento $w_{S,R}$ da matriz, corresponde ao sinal propagado pelo neurónio *R* da camada de entrada para o neurónio *S* da camada escondida), pelos valores de entrada na rede, *I*, ao qual se adiciona normalmente um viés[15] (b_S) associado a cada neurónio. A esse resultado finalmente aplica-se uma função de activação (*f*), de acordo com o problema em questão. Da aplicação da função de activação resultará um valor que será propagado para os neurónios da camada seguinte.

Existem várias funções de activação [Haykin, 1994] que são usadas conforme o tipo de rede neuronal que se está a utilizar, o intervalo de compreensão dos resultados que se pretende e obviamente do problema em questão. Neste trabalho foi utilizada uma função de activação sigmoidal. Esta função tem por domínio $\Re$ e como contradomínio o intervalo [0,1]. É ainda diferenciável, o que permite a sua utilização em redes cujo treino utilize a técnica de retropropagação[16] do erro.

[15] Em terminologia Inglesa: bias - Termo de polarização. É opcional
[16] Em terminologia Inglesa: Backpropagation



Sabendo que

$$W = [w_{S,1} \quad w_{S,2} \quad . \quad . \quad w_{S,R}]$$

$$I = [i_1 \quad i_2 \quad . \quad . \quad i_R]^T$$

então,

$$a_S = \sum_{j=1}^{R} w_{S,j} \times i_j + b_S$$

e

$$u_S = f(a_S)$$

Figura 2. Rede MLP. Processamento de informação efectuado por cada neurónio

O grande objectivo de uma rede neuronal é validar correctamente os dados de entrada, i.e. produzir as saídas mais adequadas para os exemplos que são introduzidos na rede. Para tal é necessário que todos os parâmetros da rede sejam devidamente configurados. Entre esses parâmetros incluem-se os pesos sinápticos e os parâmetros opcionais de viés. Para ajustar esses pesos para que a rede produza os melhores resultados de validação possíveis, é necessário que a rede seja devidamente treinada.

As ANNs são usualmente treinadas de uma forma supervisionada, i.e., o ajuste dos parâmetros é efectuado com base num conjunto de exemplos de treino (pares entrada, saída desejada) previamente catalogados. No processo de treino, a rede irá ajustar os seus parâmetros (*W* e *b*) de forma a que, no final, os dados de entrada sejam correctamente mapeados nos dados de saída. No caso dos problemas de classificação deste estudo, cada entrada da rede é um vector com as 40 características extraídas e cada saída desejada tem o valor 1 para o género musical correcto e zero para os restantes (Figura 3).

Neste estudo, as redes MLP são treinadas em modo *batch*, i.e., são apresentados à rede todos os pares de treino, é calculada uma medida de erro e só depois é que são ajustados os parâmetros da rede, no sentido da diminuição do erro. Na Figura 3, temos uma matriz de entrada de dimensão 40x120, na qual cada linha corresponde a uma determinada característica extraída e cada coluna corresponde ao vector de características de uma determinada música utilizada para treinar a rede. Na mesma figura está definida a matriz das saídas desejadas para a rede. Tem a dimensão 3x120 e cada coluna tem a informação sobre o género musical para a música correspondente da matriz de entrada: todas as linhas têm valor zero, excepto para a linha correspondente à classe correcta, que tem valor um. Por exemplo, se a *t-ésima* música pertencer ao género piano e o 2º neurónio de saída estiver associado à classe piano, então a *t-ésima* coluna da matriz das saídas desejadas terá o valor 1 na segunda linha e o valor zero nas restantes.

O algoritmo de treino mais utilizado em redes deste tipo é conhecido por algoritmo de retropropagação do erro [Rumelhart & McClelland, 1986; Haykin, 1994]. Neste algoritmo, o primeiro passo consiste em calcular a saída, para todas as entradas da rede. Em seguida é calculado o erro entre os valores de saída calculados e os valores de saída desejados correspondentes. Os pesos são então ajustados, da frente para trás, no sentido da redução do erro. É utilizado para tal, o método da descida do gradiente. Este processo é repetido iterativamente até que o erro esteja abaixo de um limiar definido inicialmente.

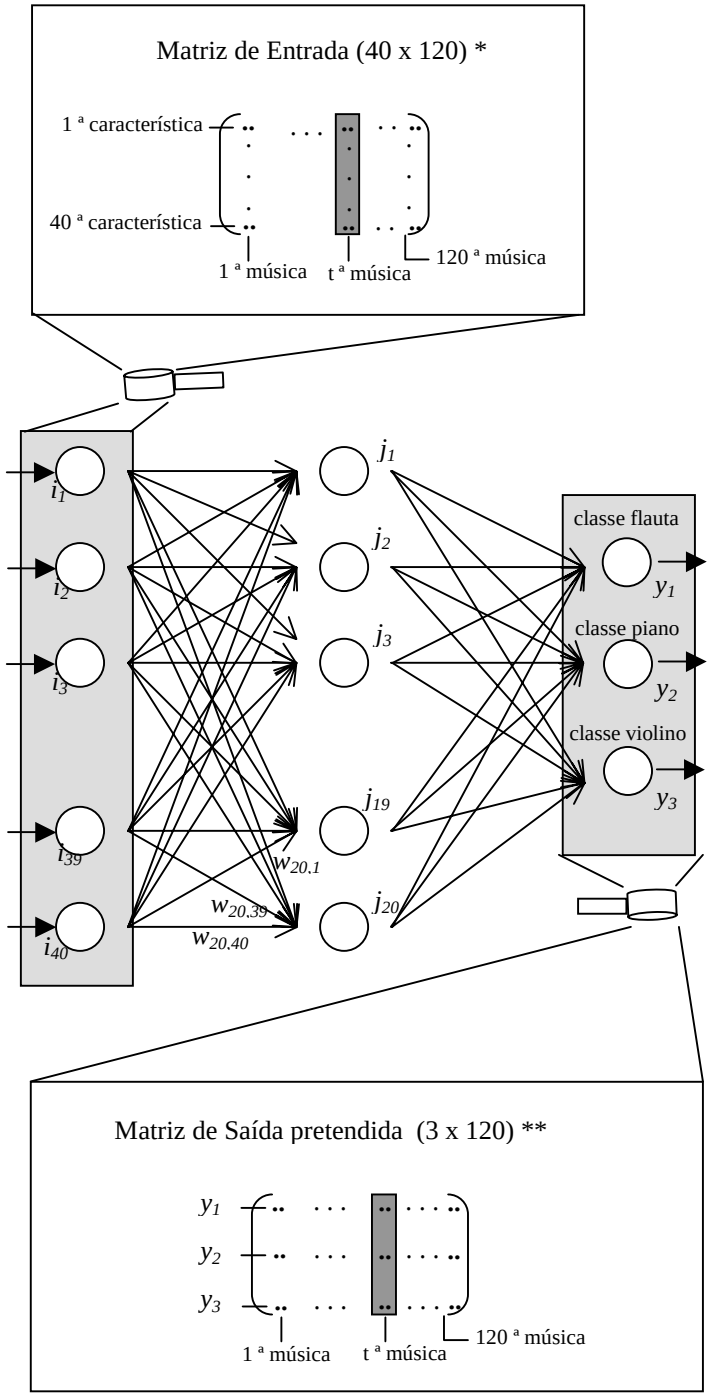


Figura 3. Rede MLP. Treino supervisionado

O método da descida do gradiente tem algumas limitações no que diz respeito às suas propriedades de convergência: o algoritmo pode convergir para um mínimo local e a selecção da velocidade de aprendizagem não é trivial (se o seu valor for muito baixo, a aprendizagem da rede será muito lenta; se for muito alto a rede poderá divergir). Existem variantes a este método que trazem algumas vantagens no sentido da

diminuição das suas limitações, e.g., aprendizagem com coeficiente de momento ou definição de uma velocidade de aprendizagem adaptativa [Haykin, 1994].

Neste estudo é utilizado o algoritmo de Levenberg-Marquardt [Hagan & Menhaj, 1994], o qual tem a vantagem de ser significativamente mais rápido (10 a 100 mais rápido [Demuth & Beale, 2001]). Além do mais, este algoritmo converge em situações onde outros não conseguem [Hagan & Menhaj, 1994].

Depois do treino, a rede neuronal tem que ser validada, i.e., a sua resposta a dados não catalogados tem de ser analisada, de forma a avaliar-se as capacidades de generalização da rede neuronal. Por conseguinte, é determinada a saída da rede correspondente a dados de entrada não catalogados. Em seguida é calculada a mesma medida de erro calculada no treino.

Os valores iniciais dos parâmetros da rede influenciam os resultados alcançados no processo de treino/validação. Como tal, este processo deve ser repetido várias vezes na tentativa de obtenção dos melhores resultados.

Tipicamente, o universo das músicas utilizadas é dividido em dois conjuntos, um para treino e outro para validação, 2/3 para o primeiro e 1/3 para o segundo respectivamente.

No sentido de evitar problemas numéricos, todas as características foram previamente normalizadas no intervalo entre 0 e 1 [Demuth & Beale, 2001].

4. Resultados Experimentais

O objectivo deste estudo é, como foi visto atrás, classificar música em cinco subgéneros da música clássica: flauta, piano, violino, coral e ópera. Estes podem ser organizados de uma forma hierárquica, como se pode visualizar na Figura 4. A taxonomia ilustrada na figura foi definida apenas por uma questão de clareza. Na prática a classificação efectuada não foi hierárquica.

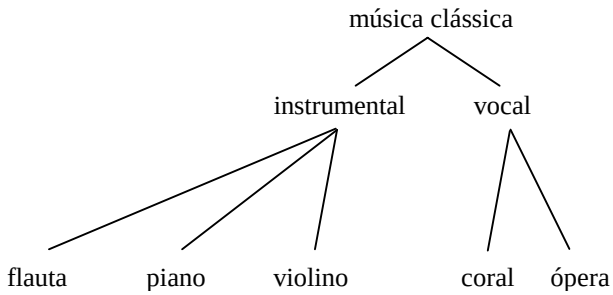


Figura 4. **Taxonomia utilizada**

O primeiro problema de classificação consiste na discriminação entre três géneros de música instrumental: música para flauta, para piano e para violino. Assim, uma peça musical pertence a cada uma das três classes se essa peça for um solo de um desses instrumentos (e.g., um ou mais violinos) ou, tendo orquestra, esses instrumentos forem predominantes.

No segundo problema, o objectivo é classificar em subgéneros da música vocal: música coral e ópera. Antes de mais nada, importa definir o que se entende por cada um destes estilos, já que nomeadamente a ópera é de facto uma representação teatral em que os actores cantam poemas líricos com o acompanhamento de uma orquestra. Há uma característica que a distingue normalmente, que tem a ver com o tipo de solistas normalmente utilizados (e.g. tenor[17], soprano[18], mezzo-soprano[19]). As técnicas de canto utilizadas na ópera (e.g., *tremolo*, *vibrato*), faz com que normalmente o ser humano consiga distinguir com uma percentagem razoável de sucesso a ópera da música coral. Por sua vez, este último estilo é normalmente caracterizado por várias vozes em coro, sem *vibrato* nem *tremolo*. É ainda usual que as peças de música coral sejam *a cappela*, i.e. sem parte instrumental.

Finalmente, o terceiro problema de classificação consiste na discriminação entre os cinco géneros musicais anteriores.

Foram utilizados para os três problemas de classificação redes MLP com três camadas. Estas redes neuronais foram treinadas em modo *batch* com o algoritmo de Levenberg-Marquardt. Cada rede é constituída por 40 neurónios na camada de entrada (um por cada característica extraída), um número variável de neurónios na camada escondida (descrição abaixo) e 2, 3 ou 5 neurónios na camada de saída conforme o

[17] Tenor é a voz masculina mais alta. Enrico Caruso, Ben Heppner e Luciano Pavarotti são três famosos tenores.
[18] Soprano é a voz feminina mais alta. Joan Sutherland e Maria Callas são duas sopranos famosas sopranos.
[19] Mezzo-soprano é a voz feminina entre o soprano e o contralto. Cecilia Bartoli, Marilyn Horne e Anne Sofie Von Otter são três famosas mezzo-sopranos.

problema de classificação em estudo. Tanto a camada escondida como a camada de saída utilizam funções de activação sigmoidais. Foram utilizadas para treino um total de 40 peças musicais de cada género e para validação 20.

A validação, i.e., a classificação de peças desconhecidas, foi efectuada segundo duas perspectivas que serão identificadas deste ponto em diante, por regra de cálculo de percentagem 1 (RCP1) e regra de cálculo de percentagem 2 (RCP2).

Irão ser descritas em seguida as duas perspectivas anteriores.

Regras de cálculo de percentagens 1

Considera-se que uma música de um determinado género musical é bem classificada se o valor mais alto de saída da rede pertence a esse género e esse valor é maior ou igual a 0,7 (de recordar que os valores de saída da rede variam entre 0 e 1). Quando uma música é bem classificada, é-o sem margem para dúvidas.

Quando todos os valores de saída da rede são inferiores a 0,7, a música é considerada sem classificação. O valor mais alto não é suficientemente alto para evitar ambiguidades na classificação.

É ainda calculado, para permitir distinguir melhor os resultados obtidos para as várias classes, um campo representado por *{gn 2 <= 0,2}*. O objectivo deste campo, é apresentar de entre as músicas bem classificadas, aquelas cuja distância ao 2º género com valor mais alto, é inferior ou igual a 0,2 (e.g., para uma música bem classificada, o valor mais alto de saída da rede é 0,8 e o 2º valor mais alto é 0,65; esta situação está no âmbito de aplicação desta regra). Uma peça musical nesta situação é apenas predominantemente de um dado género musical.

Se nenhuma das regras anteriores for aplicada, conclui-se que a música em questão foi mal classificada.

Regras de cálculo de percentagens 2

Nesta perspectiva, considera-se que uma música de um dado género é bem classificada se o valor de saída da rede para esse género for o mais alto, independentemente da ordem de grandeza.

Se a regra anterior não for aplicada, então é porque a música foi mal classificada.

Considera-se ainda um campo representado por *{gn 2 >= 0,7}*. Este campo contabiliza, de entre as músicas mal classificadas, aquelas cujo 2º valor mais alto de saída da rede corresponde ao género correcto e além disso esse valor é maior ou igual a 0,7. (e.g., uma música obtém 0,96 para uma saída da rede que não corresponde ao género correcto e 0,75 para a saída que corresponde ao género correcto; então, essa música será contabilizada no âmbito desta regra). Esta regra permite detectar as músicas que, embora mal classificadas, tenham valores suficientemente altos para o género a que pertencem.

Primeira Classificação: Três Géneros

Neste primeiro problema, as peças musicais foram classificadas em três classes: flauta, piano e violino.

Para determinar o número mais adequado de neurónios para a camada escondida, foram testados vários valores no intervalo [10, 30]. Os melhores resultados de classificação foram obtidos com 20 neurónios. Esses resultados reflectem-se por uma taxa de sucesso de 83,3% para RCP1 e de 85% para RCP2.

As Tabelas 1 e 2 apresentam, respectivamente para RCP1 e RCP2, os resultados para o melhor conjunto treino/validação:

RCP1 83,3%	flauta	piano	violino
flauta	85	10	5
piano	5	80	10
violino	5	10	85
s/classificação	5	0	0
gn 2 < = 0,2	0	0	0

Tabela 1. Matriz de confusão da música instrumental, RCP1: MLP.

RCP2 85%	flauta	piano	violino
flauta	90	10	5
piano	5	80	10
violino	5	10	85
gn 2 >= 0,7	0	0	10

Tabela 2. Matriz de confusão da música instrumental, RCP2: MLP.

As colunas da matriz representam o género a que pertence a música e as linhas o resultado da validação. Assim, por exemplo para RCP1 (Tabela 1), 80% das músicas de piano foram bem classificadas, 5% das músicas de violino foram classificadas como flauta e 5% das músicas de flauta não obtiveram classificação.

Analisando RCP1 (Tabela 1), conclui-se que foram obtidas como percentagem de músicas bem classificadas, 85% para flauta, 80% para piano e 85% para violino. Em relação, por exemplo, às peças de flauta, 5% foram mal classificadas como piano, 5% foram classificadas como violino e 5% não obtiveram classificação (verdadeiros negativos). Verifica-se também que 10% das peças para piano e 5% das peças para violino são erradamente classificadas como sendo peças para flauta (falsos positivos). Infere-se ainda que a distância entre o valor do género correcto e o segundo valor mais alto é sempre superior a 0,2 ("*gn 2 < 0,2*" = 0%), logo as músicas bem classificadas, são-no sem margem para dúvidas.

Em relação aos dados referentes a RCP2 (Tabela 2), a percentagem de músicas correctamente classificadas foi de 90% para flauta, 80% para piano e 85% para violino. Conclui-se ainda que 10% das músicas de violino que foram mal classificadas, obtiveram para a classe violino o segundo valor mais alto e esse valor é superior a 0,7 ("*gn 2 > 0,7*" = 10%).

Analisando os erros de classificação, reparou-se que aconteceram essencialmente em músicas nas quais os instrumentos são tocados de uma forma pouco usual para esses mesmos instrumentos. Com certeza esse tipo de músicas tem uma menor representação nos exemplos de treino da rede. Por exemplo, há 2 músicas de violino que obtiveram valores superiores a 0,7 para a classe violino, mas foram ultrapassados pelos valores da classe piano – 2 extractos musicais de *Bach* e *Mozart*. Esses extractos têm em comum o facto de os instrumentos principais (violinos) serem tocados de uma forma muito lenta e terem poucas variações de amplitude. Essas características são típicas das músicas para piano. Os valores bastante altos para a classe violino explicam-se por, apesar de em diversos aspectos não serem músicas típicas de violino, as características tímbricas extraídas de cada uma dessas músicas detectarem a presença desses instrumentos.

Segunda Classificação: Dois Géneros

Nesta situação, as peças musicais foram classificadas em dois estilos: ópera e música coral. Os melhores resultados de classificação foram obtidos com 25 neurónios na

camada escondida: uma percentagem média de sucesso na classificação de 90% tanto para RCP1 como para RCP2.

As tabelas 3 e 4 resumem a classificação para o melhor conjunto treino/validação deste problema.

RCP1 90%	coral	ópera
coral	90	10
ópera	10	90
s/classificação	0	0
gn 2 <= 0,2	0	0

Tabela 3. Matriz de confusão da música vocal, RCP1: MLP.

RCP2 90%	coral	ópera
coral	90	10
ópera	10	90
gn 2 >= 0,7	0	0

Tabela 4. Matriz de confusão da música vocal, RCP2: MLP.

Analisando RCP1 (Tabela 3), conclui-se que foram obtidas como percentagem de músicas bem classificadas, 90% para música coral e 90% para ópera.

De notar que as percentagens anteriores mostram total ausência de ambiguidade nesta classificação, já que as distâncias entre as músicas correctamente classificadas e o segundo valor mais alto são superiores a 0,2 ("*gn 2 <= 0,2*" = 0%).

Quanto a RCP2 (Tabela 4), os resultados obtidos são os mesmos de RCP1, i.e., 90% de músicas bem classificadas, tanto para música coral como para ópera.

São quatro as peças musicais mal classificadas, duas de ópera e duas corais. Em relação às músicas corais que são erradamente classificadas como peças de ópera, uma delas tem parte instrumental, ao contrário da maioria das músicas de treino dessa classe que são *a cappela*. Essa música tem ainda a sobressair uma voz feminina cuja prestação pode facilmente, para a média dos humanos ser confundida com ópera. A outra música coral tem várias vozes cujas prestações atingem frequências altas em todo o excerto. Quanto aos excertos de ópera mal classificados, tirando o facto de serem um pouco atípicos em relação a essa classe, já que são partes bastante calmas de ópera e com pausas, sendo mesmo uma delas *a cappela* (como a maioria das peças corais), não se encontram razões claras para o erro na classificação. A única conclusão que se pode

tirar, é que talvez as características extraídas das peças musicais sejam bastante boas para os casos bem comportados, sendo necessário a inclusão de novas características e/ou a eliminação de características redundantes para os casos mais atípicos.

Terceira Classificação: Cinco Géneros

O objectivo deste problema é catalogar música numa das cinco classes anteriores: flauta, piano, violino, coral e ópera. Os melhores resultados de classificação foram de 64% para RCP1 (20 neurónios na camada escondida) e 76% para RCP2 (30 neurónios na camada escondida).

A Tabela 5 apresenta os resultados de RCP1 para a rede escolhida.

RCP1 64%	flauta	piano	violino	coral	ópera
flauta	65	15	5	0	10
piano	10	65	0	10	0
violino	0	10	70	10	0
coral	15	0	5	50	0
ópera	0	0	5	15	70
s/classificação	10	10	15	15	20
gn 2 <= 0,2	0	10	20	5	15

Tabela 5. Matriz de confusão das músicas instrumental e vocal, RCP1: MLP.

Como se pode verificar pela da tabela anterior, foram obtidas como percentagens de músicas correctamente classificadas, 65% para flauta, 65% para piano, 70% para violino, 50% para coral e 70% para ópera. A percentagem geral de sucesso foi de 64%.

Os resultados de validação para RCP2 são ilustrados na tabela seguinte (Tabela 6).

RCP2 76%	flauta	piano	violino	coral	ópera
flauta	75	20	0	10	10
piano	5	65	0	15	5
violino	0	5	85	0	0
coral	10	5	10	75	5
ópera	10	5	5	0	80
gn 2 >= 0,7	0	5	5	0	0

Tabela 6. Matriz de confusão das músicas instrumental e vocal, RCP2: MLP.

Segundo RCP2, a percentagem de sucesso na classificação foi de 75% para flauta, 65% para piano, 85% para violino, 75% para coral e 80% para ópera. Nota-se aqui que tendo em conta a junção de géneros musicais tão distintos, uma classificação global de 76% pode ser considerada bastante razoável.

A análise das duas tabelas anteriores parece indicar que a aprendizagem das características de cada género musical foi superior nesta última rede (Tabela 6). Para tentar comprovar isso, vai ser apresentada em seguida a classificação RCP1 correspondente à rede com 30 neurónios na camada escondida (Tabela 7).

RCP1 62%	flauta	piano	violino	coral	ópera
flauta	65	20	0	5	0
piano	0	65	0	0	0
violino	0	5	70	0	0
coral	5	5	5	50	5
ópera	0	0	5	0	60
s/classificação	30	5	20	45	35
gn 2 <= 0,2	10	0	0	15	5

Tabela 7. Matriz de confusão das músicas instrumental e vocal, RCP1 (2): MLP.

Como se verifica na tabela anterior, as percentagens de músicas bem classificadas são muito parecidas com as da rede melhor para RCP1 (Tabela 5) e as classificações globais são muito próximas: 64% e 62%.

O que se nota na Tabela 7 em relação à Tabela 5 é uma muito menor percentagem média de músicas mal classificadas: 11% contra 22%. Em contrapartida na Tabela 7 há uma percentagem média muito maior de músicas que não obtiveram qualquer classificação: 27% contra 14%. A juntar a isto, ainda o facto de o campo *{gn 2 <= 0,2}* ter tido um valor médio mais baixo em (7) que em (5): 6% contra 10%. Estes resultados, no seu conjunto, mostram que o classificador considerado em (7) aprendeu com maior precisão as características fundamentais de cada género musical, já que tendo uma percentagem de sucesso muito parecido com o classificador de (5) tem muito menos músicas mal classificadas, preferindo antes não lhes atribuir qualquer classificação. Além do mais há uma percentagem menor de músicas que apesar de bem classificadas o foram com alguma ambiguidade, como se pode atestar pela diferença de percentagens de *{gn 2 <= 0,2}*: 6% contra 10%. É evidente ainda que para um sistema real de

classificação seria sempre preferível, apesar de tudo uma não classificação que uma má classificação.

No caso deste terceiro problema de classificação em cinco géneros musicais esperava-se sem dúvida resultados menos precisos do que nos primeiro e segundo problemas, já que nesses os géneros, além de serem em menor número, são mais parecidos e têm sempre algo em comum (instrumentos apenas e voz). As respectivas redes são treinadas fundamentalmente no sentido da distinção do timbre e do *pitch* dos instrumentos principais e da voz.

No terceiro problema misturam-se géneros instrumentais com géneros vocais, numa rede que não é treinada explicitamente para distinguir música instrumental de música vocal. Depois de escutados novamente alguns dos extractos de música mal classificados, concluiu-se que muitos deles foram confundidos com outros géneros pela forma como eram interpretados.

5. Sistema de Classificação Automática

Nesta secção, pretende-se fazer uma aproximação a um sistema real de classificação automática de géneros musicais. Este sistema irá tentar validar correctamente um conjunto de 100 músicas, 20 de cada uma das classes, flauta, piano, violino, coral e ópera. De cada música são extraídos 10 extractos de 6s, escolhidos de igual forma para todas as músicas. Cada música será classificada no género musical mais representado entre os seus 10 extractos.

De notar que as 100 músicas a validar são as correspondentes ao conjunto de validação inicial. A diferença reside no facto de que, inicialmente foi extraído um extracto teoricamente “bem comportado”, enquanto que aqui são extraídos 10 extractos de uma forma perfeitamente automática, em que o critério [Malheiro, 2004] de escolha dos extractos está relacionado apenas com a sua posição relativa na música.

As regras básicas de classificação continuam a ser as aplicadas na secção anterior, ou sejam, RCP1 e RCP2. No entanto foram definidas mais algumas regras relacionadas especificamente com a validação das músicas baseada na validação dos seus extractos.

Assim, para RCP1, uma música é classificada no género musical mais representado pelos seus extractos. Se existirem mais extractos sem classificação do que classificados num determinado género musical, então a música será considerada sem classificação. Se

dois ou mais géneros estiverem igualmente representados e não existir outro género com maior representação, então a música será classificada nesses géneros. Se existir igual representação de um ou mais géneros e de extractos sem classificação e não existir outro género com maior representação, então a música será classificada nesses géneros.

Para RCP2, uma música é classificada no género musical mais representado pelos seus extractos. Se existir igual representação de dois ou mais géneros, a música pertencerá ao género ou géneros que em RCP1 tenham tido maior representação. Caso se mantenham igualmente representados então a música será classificada nesses géneros, (e.g., em RCP1, coral=4, ópera=5, s/classificação=1. Em RCP2, coral=ópera=5. A música é classificada em ópera, segundo RCP1 e RCP2).

Serão visualizados em seguida, através de matrizes de confusão de géneros musicais, os resultados alcançados pelos classificadores em cada um dos problemas de classificação. As informações pormenorizadas sobre a classificação de cada um dos extractos podem ser consultadas em [Malheiro, 2004].

Primeira Problema: Três Géneros

Neste primeiro problema pretende-se classificar um total de 60 músicas, 20 de cada um dos estilos: flauta, piano e violino. Cada música é representada por 10 extractos e a sua classificação resulta no género musical mais representado pelos seus extractos.

Vão ser resumidos nas tabelas seguintes os resultados alcançados pelas 60 músicas de validação, (Tabela 8 e Tabela 9).

RCP1 78%	flauta	piano	violino
flauta	75	9	-
piano	5	59,1	-
violino	20	31,9	100
s/classificação	-	-	-

Tabela 8. Matriz de confusão do protótipo para a primeira classificação: RCP1.

RCP2 78%	flauta	piano	violino
flauta	75	9	-
piano	5	59,1	-
violino	20	31,9	100

Tabela 9. Matriz de confusão do protótipo para a primeira classificação: RCP2.

A percentagem de músicas bem classificadas foi de 78%, independentemente das perspectivas de classificação. Por género, foram bem classificadas, 75% das músicas de flauta, 59,1% das de piano e 100% das de violino. Não houve nenhuma música sem classificação.

De notar os resultados excelentes para a classe violino que obteve 100%. Todas as músicas foram categoricamente classificadas, o que mostra que a rede aprendeu a identificar correctamente as características do instrumento violino, nomeadamente o seu timbre.

As classificações de piano desiludiram um pouco com apenas 59,1% de músicas correctamente classificadas. 31,9% foram classificadas erradamente na classe violino. Não se encontrou uma justificação razoável para este facto.

De qualquer forma estes resultados são considerados promissores, já que os resultados globais alcançados com esta extracção cega de peças (78%) ficaram bastante perto da extracção de exemplos “bem comportados” (85%).

Segundo Problema: Dois Géneros

O objectivo deste problema é a classificação de música num de dois géneros musicais: coral e ópera. Vão ser classificados 400 extractos musicais, pertencentes a um total de 40 músicas. Dessas, 20 são corais e 20 são óperas.

Os resultados da análise da tabela anterior vão ser resumidos em seguida nas Tabelas 10, para RCP1, e 11, para RCP2.

RCP1 73,5%	coral	ópera
coral	81,8	34,8
ópera	18,2	65,2
s/classificação	-	-

Tabela 10. Matriz de confusão do protótipo para a segunda classificação: RCP1.

RCP2 73,5%	coral	ópera
coral	81,8	34,8
ópera	18,2	65,2

Tabela 11. Matriz de confusão do protótipo para a segunda classificação: RCP2.

A percentagem de músicas bem classificadas neste segundo problema de classificação foi de 73,5% tanto para RCP1 como para RCP2. Esta percentagem resulta da média das percentagens por género, que foram de 81,8% para coral e de 65,2% para ópera.

Como se pode visualizar na Tabela 10, nenhuma música foi considerada sem classificação.

Estes resultados surpreenderam um pouco pela negativa, já que houve uma grande descida na percentagem global de sucesso entre a classificação com exemplos "bem comportados" e a classificação cega. Essa descida foi de 90% para 73,5% e está fundamentalmente relacionada com a percentagem grande de músicas de ópera que foram erradamente classificadas na classe coral (34,8%). Analisando alguns desses casos, constatou-se que algumas partes da maioria das óperas têm grandes parecenças com a música coral, principalmente nas partes mais calmas. Na classificação que obteve melhores resultados, foram utilizados essencialmente exemplos típicos de ópera.

Terceiro Problema: Cinco Géneros

Neste último problema pretende-se fazer a classificação de 100 músicas por cinco géneros musicais. Essa classificação depende, para cada música, da classificação dos seus 10 extractos. O género mais representado nos extractos será considerado o género da música.

Os resultados da análise da tabela anterior vão ser resumidos em seguida nas tabelas (12) para RCP1 e (13) para RCP2.

RCP1 57,3%	flauta	piano	violino	coral	ópera
flauta	59,2	3,9	0	18,2	0
piano	4,5	42,3	0	4,5	0
violino	13,6	7,7	85	0	9,1
coral	4,5	19,2	5	59,2	31,8
ópera	0	11,5	0	4,5	40,9
s/classificação	18,2	15,4	10	13,6	18,2

Tabela 12. Matriz de confusão do protótipo para a terceira classificação: RCP1.

Verifica-se pela tabela anterior que a percentagem de sucesso de músicas bem classificadas foi, respectivamente para flauta, piano, violino, coral e ópera, de 59,2%, 42,3%, 85%, 59,2% e 40,9%. A taxa geral de sucesso foi de 57,3%. A percentagem de músicas sem classificação foi de 15%.

RCP2 66,7%	flauta	piano	violino	coral	Ópera
flauta	66,7	4,2	0	19,1	5
piano	4,8	50	0	4,7	0
violino	14,2	4,2	100	0	15
coral	9,5	29,1	0	66,7	30
ópera	4,8	12,5	0	9,5	50

Tabela 13. Matriz de confusão do protótipo para a terceira classificação: RCP2.

Através da tabela anterior, verifica-se que a percentagem de músicas bem classificadas foi de 66,7% para flauta, 50% para piano, 100% para violino, 66,7% para coral e 50% para ópera. A percentagem geral de sucesso foi de 66,7%.

De notar, antes de mais, a classificação conseguida pela classe violino. Segundo RCP2, todas as suas músicas foram correctamente classificadas. Em RCP1, existiram apenas 3 músicas que não foram classificadas em violino: duas sem classificação e uma classificada em coral. Conclui-se que o classificador aprendeu da melhor forma a identificar as características do instrumento violino.

Em oposição à classe violino, as classes piano e ópera obtiveram classificações que desiludiram. Através da inspecção à Tabela 13 repara-se que 29,1% das músicas de piano foram classificadas em coral e 30% das de ópera foram classificadas em coral. Esta confusão entre ópera e coral já tinha sido detectada no segundo problema de classificação. Por análise de alguns casos, detectou-se que partes atípicas da ópera,

muito lentas e com interpretações parecidas com a música coral, são classificadas facilmente em coral. Isto está relacionado com o facto dos classificadores terem sido treinados essencialmente com casos típicos de ópera. Além disso a ópera não mantém normalmente durante toda a sua duração as características que a identificam facilmente. Uma solução para aumentar a fiabilidade da classificação pode passar numa primeira fase por aumentar o número de extractos de cada música. Outra possível solução será treinar a rede com um conjunto maior de exemplos de treino que contenha mais casos atípicos.

Quanto ao facto da confusão entre piano e coral, não se encontrou uma justificação razoável. No entanto há parecenças entre os casos típicos dos dois estilos no que toca à sonoridade: ambos são bastante calmos.

Pode-se considerar que os resultados alcançados por esta abordagem a um sistema real de classificação são no mínimo interessantes numa perspectiva de melhoria futura, já que foi obtida uma taxa de sucesso de 66,7% relativamente próxima da taxa que tinha sido obtida para casos mais típicos de classificação (76%).

6. Conclusões

O objectivo principal deste artigo foi apresentar uma metodologia para a classificação automática de música clássica. Apesar dos resultados alcançados não serem ainda os suficientes para as exigências de uma aplicação real, são bastante promissores.

No caso mais complexo de classificação, onde foram definidas cinco categorias, os resultados não foram tão bons. No entanto, na nossa opinião, um classificador hierárquico seguindo a estrutura da Figura 4, produziria melhores resultados.

No futuro iremos efectuar uma análise mais detalhada do espaço de características, nomedamente no que diz respeito à detecção e eliminação de características redundantes, bem como à definição e utilização de novas características que possam ajudar a discriminar os casos mais atípicos. Adicionalmente, pretende-se fazer a expansão da árvore em profundidade e em largura, i.e., mais classes e subclasses. No caso da utilização de categorias como a valsa, características rítmicas, não utilizadas aqui, serão com certeza importantes.

Referências

[Broomhead & Lowe, 1988]

Broomhead D.S. & Lowe D., 1988, "Multivariable function interpolation and adaptativo networks", *Complex Systems*, Vol. 2, pp 321-355.

[Demuth & Beale, 2001]

Demuth, H. & Beale, M., 2001, "*Neural Network Toolbox User's Guide*", version 4, Mathworks.

[Fraser & Fujinaga, 1999]

Fraser, A. & Fujinaga, I., 1999, "Toward real-time recognition of acoustic musical instruments", *Proceedings of the International Computer Music Conference - ICMC 1999*.

[Golub, 2000]

Golub, S., 2000, "*Classifying Recorded Music*", MSc Thesis, University of Edinburgh.

[Hagan & Menhaj, 1994]

Hagan, M. & Menhaj, M., 1994, "Training Feedforward Networks with the Marquardt Algorithm", *IEEE Transactions on Neural Networks*, vol. 5, no. 6, pp. 989-993.

[Haykin, 1994]

Haykin S., 1994, "*Neural Networks: A Comprehensive Foundation*", Macmillan College Publishing.

[Kohonen, 1989]

Kohonen T., 1989, "*Self-Organization and Associative Memory*", 3rd edition, Springer-Verlag, Berlin.

[Kosina, 2002]

Kosina, K., 2002, "*Music Genre Recognition*", MSc Thesis, Hagenberg.

[Malheiro, 2004]

Malheiro, R., 2004, "*Sistemas de Classificação Automática em Géneros Musicais*", MSc Thesis, Universidade de Coimbra, em português.

[Martin, 1998]

Martin, K., 1998, “Toward Automatic Sound Source Recognition: Identifying Musical Instruments”, *NATO Computational Hearing Advanced Study Institute*, Il Ciocco, Italy.

[Martin & Kim, 1998]

Martin, K. & Kim, Y., 1998, “Musical instrument identification: A pattern-recognition approach”, *Proceedings of the 136th meeting of the Acoustical Society of America - ASA 1998.*

[Perrot e Gjerdigen, 1999]

Perrot D. & Gjerdigen R., 1999, “Scanning the dial: An exploration of factors in identification of musical style”, *Society for Music Perception and Cognition*, pp. 88, 1999.

[Polikar, 2003]

Polikar, R., 2003, “*The Wavelet Tutorial*”, http://engineering.rowan.edu/ ~polikar/WAVELETS/WTtutorial.html, disponível em Julho de 2003.

[Rumelhart & McClelland, 1986]

Rumelhart, D. & McClelland, J., 1986, “*Parallel Distributed Processing, Explorations in the Microstructure of Cognition*”, Vol. 1:Foundations, MIT Press, Cambridge, USA.

[Sarle, 2001]

Sarle W. (maintainer), 2001, “*Neural Nets FAQ*”, ftp://ftp.sas.com/pub/neural/ FAQ3.html.

[Smith, 1997]

Smith, S, 1997, “*The Scientist and Engineer’s Guide to Digital Signal Processing*”, California Technical Publishing.

[Tzanetakis & Cook, 2002]

Tzanetakis G. & Cook P., 2002, “Musical Genre Classification of Audio Signals”, *IEEE Transactions on Speech and Audio Processing*, 10(5), 293-3 02.

[Tzanetakis et al., 2001]

Tzanetakis, G., Essl, G. & Cook, P., 2001, “Automatic Musical Genre Classification of Audio Signals”, *Proceedings of International Symposium on Music Information Retrieval - ISMIR 2001*.

[Wold et al., 1996]

Wold, E., Blum, T., Keislar, T. & Wheaton, J., 1996, “Content-based classification, search and retrieval of audio”, *IEEE Multimedia*, Vol. 3, nº 2.