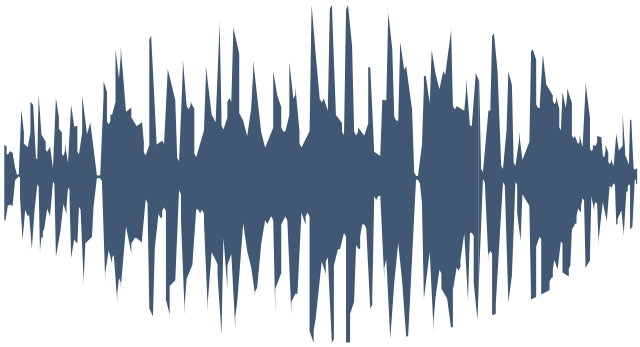
1 2 90

UNIVERSIDADE D COIMBRA



TRABALHO PRÁTICO 2

MUSIC INFORMATION RETRIEVAL

Ana Carolina Morais Nº2021222056
Fernanda Fernandes Nº2021216620
Miguel Miranda Nº2021212100
Rodrigo Sá Nº2021213188

INDICE

INTRODUÇÃO

Este trabalho prático tem como principal objetivo adquirir sensibilidade para as questões fundamentais dos MMIR, em particular, à extração de informação presente em áudio e, com esta, desenvolver um sistema de recomendação de músicas a um utilizador.

Uma vez, que a tecnologia dos MMIR abrange uma ampla gama de aplicações, incluindo a recomendação de música, a classificação automática de género, identificação de artistas e instrumentos, transcrição de áudio, separação de fontes sonoras, entre outras.

Este projeto inclui tarefas como a extração de características de áudio de uma base de dados de 900 músicas, a implementação de métricas de similaridade, e a avaliação objetiva e subjetiva das recomendações de música geradas.

Assim, todo o código foi desenvolvido na linguagem de programação *Python*, com recurso às bibliotecas *librosa, numpy, scipy*, *os*, entre outras.

EXERCÍCIO 1

• PREPARAÇÃO

Na fase de preparação, é feito o download de uma base de dados do "***4Q Audio Emotion Dataset***", que contém 900 músicas. Após o *download* analisam se os arquivos de áudio, que estão no formato mp3, com especial atenção na query fornecida e nos metadados presentes no arquivo *panda_dataset_taffc_metadata.csv*. Além disso, é necessário o estudo da framework de processamento de áudio *librosa*, que posteriormente será necessária a sua instalação.

O código fonte de base fornecido no arquivo *mrs.py* é executado e analisado, sendo necessário mais uma vez posteriormente a instalação da biblioteca *sounddevice* para reprodução de arquivos de áudio. Por fim, é necessária a leitura e estudo da documentação da *librosa*, especialmente das funções relacionadas com a extração de características de áudio, disponível online.

EXERCÍCIO 2

• EXTRAÇÃO DE *FEATURES*

Nesta parte inicial do projeto, o objetivo é extrair dados sobre um conjunto de 900 músicas fornecidas no dataset do trabalho e usar a biblioteca *librosa* para obter, no total, 190 features.

Os dados obtidos através do uso da *librosa*, encontram-se representados numa tabela, abaixo.

mfcc	centroid	bandwidth	constrast	flatness	rolloff	F0	rms	Zero crossing rate	tempo

Tabela 1 – Features extraídas com o auxílio do *librosa*.

É de realçar que quase todas as *features* contidas acima, são representadas por um array 1x 1295, excepto a feature tempo, que é um valor único, o mffc representado por 13x1285 (os 13 níveis mais importantes) e o spectral contrast, 7x1295.

À medida que os dados são extraídos, começam a ser calculadas estatísticas que serão utilizadas para testes e recomendações futuras. São sete as estatísticas computadas: média, desvio padrão, assimetria, *curtose*, mediana, valor máximo e valor mínimo. Uma vez concluído este processo, o resultado é uma matriz de 900 por 190, contendo as sete estatísticas aplicadas a cada vetor unidimensional extraído de cada música. O tempo, por sua vez, é armazenado na matriz final como um valor único, exatamente como foi extraído.

Após a conclusão do processo de normalização, os dados são salvos num ficheiro, denominado "featuresNormalized.csv". A importância de armazenar estas informações decorre do tempo necessário para extrair todos os dados utilizando a biblioteca *Librosa*, que pode ser extenso. Portanto, se o ficheiro já tiver sido gerado anteriormente, é possível simplesmente aceder as informações nele contidas, evitando a repetição de um processo demorado. Isso resulta numa economia significativa de tempo de execução e permite a obtenção dos resultados do trabalho de forma muito mais eficiente. Esta verificação, da existência dos ficheiros, está implementada ao longo do projeto todo.

IMPLEMENTAÇÃO SPECTRAL CENTROID

Desenvolvemos uma implementação própria para calcular o *spectral centroid* de um sinal usando Python com as bibliotecas Numpy e Scipy. A função my_spectral_centroid emprega a função np.fft.rfft para obter o espectro de frequência de cada janela do sinal, aplicando uma janela de Hanning para minimizar o vazamento espectral.

Os centroides espectrais são então calculados para cada frame, ponderando as frequências pelo seu respetivo espectro e normalizando pela soma das magnitudes. Por fim, comparamos os resultados desta implementação com os dados obtidos pela função de *spectral centroid* da *Librosa*, utilizando o coeficiente de correlação de Pearson e o *RMSE* como métricas de comparação. A análise revelou que é necessário ajustar o índice inicial dos dados extraídos pela *Librosa* para sincronizar com os resultados do código implementado, devido a um atraso de duas janelas na extração de *features* pela mesma.

EXERCÍCIO 3

Para este exercício, implementamos três métricas de similaridade distintas, para calcular as distâncias entre características normalizadas de músicas. As métricas implementadas incluem a distância Euclidiana, a distância de Manhattan e a distância do Coseno. Cada uma dessas métricas fornece uma perspectiva diferente sobre a similaridade entre duas instâncias de dados, sendo essenciais para sistemas de recomendação e outras aplicações de recuperação de informação.

Para o cálculo das distâncias, utilizámos as fórmulas representadas abaixo, sendo que no caso da distância do coseno, utilizámos o método *cosine()* da biblioteca *scipy*.

$$d(F_1, F_2) = \sqrt{\sum_{i=1}^{N} (F_1(i) - F_2(i))^2}$$

Distância euclidiana

$$d(F_1, F_2) = \sum_{i=1}^{N} |F_1(i) - F_2(i)|$$

Distância de Manhattan

$$d(F_1, F_2) = 1 - \frac{\sum_{i=1}^{N} F_1(i) \cdot F_2(i)}{\sqrt{\sum_{i=1}^{N} (F_1(i))^2} \cdot \sqrt{\sum_{i=1}^{N} (F_2(i))^2}}$$

Distância de Coseno

Posteriormente, calculamos as três métricas de distância para cada música em relação à *query* normalizada. Os resultados são armazenados em três ficheiros CSV distintos: "euclideanDistance.csv", "manhattanDistance.csv" e "cosineDistance.csv".

Para a *query* especificada, geramos três 'rankings' de similaridade, cada um correspondendo a uma métrica de distância. Cada 'ranking' lista as 10 músicas mais similares conforme determinado pela métrica correspondente. Obtemos assim os seguintes resultados:

RANKING

Ranking Euclidiana	Query
Música 1	MT0000414517
Música 2	MT0004274911
Música 3	MT0003949060
Música 4	MT0000218346
Música 5	MT0001624303
Música 6	MT0003900455
Música 7	MT0004032071
Música 8	MT0009208842
Música 9	MT0001515531
Música 10	MT0005752234

Ranking Manhattan	Query
Música 1	MT0000414517
Música 2	MT0003949060
Música 3	MT0000218346
Música 4	MT0004274911
Música 5	MT0001624303
Música 6	MT0000040632
Música 7	MT0034125967
Música 8	MT0003900455
Música 9	MT0005469880
Música 10	MT0009208842

Ranking Cosine	Query
Música 1	MT0000414517
Música 2	MT0004274911
Música 3	MT0003949060
Música 4	MT0000218346
Música 5	MT0001942272
Música 6	MT0002634024
Música 7	MT0009208842
Música 8	MT0004032071
Música 9	MT0003900455
Música 10	MT0001624303

Analisando as tabelas obtidas, conseguimos perceber uma semelhança entre os ‘rankings’ de similaridade das 3 distâncias, verificando que os ‘rankings’ são quase inteiramente compostos pelas mesmas músicas, apenas variando a posição em que as mesmas são apresentadas.

EXERCÍCIO 4

• AVALIAÇÃO OBJETIVA

Para este último exercício, estamos perante a parte da avaliação, sendo a primeira componente objetiva. O objetivo é extrair os metadados do ficheiro "panda_dataset_traffc_metadata.csv" e, com base nestes, percorrer todas as músicas e assim calcular o quão próxima cada uma está de cada música presente na *query*.

Música	**MT0000414517**
Música 1	MT0033397838
Música 2	MT0027048677
Música 3	MT0000040632
Música 4	MT0012331779
Música 5	MT0003949060
Música 6	MT0010487769
Música 7	MT0010489498
Música 8	MT0002222957
Música 9	MT0008222676
Música 10	MT0007840454

O sistema de 'ranking' beneficia as similaridades encontradas, ou seja, por cada característica em comum (artista, género, quadrante e emoção), é atribuído um ponto. No final, é gerado um 'ranking' final de similaridade, no qual constam as 10 músicas, que obterem melhor pontuação para cada uma das músicas que constam na *query* previamente explicada.

Assim, encontra-se ao lado a tabela com o 'ranking'.

A seguir, irar-se-à calcular a precisão de cada escolha feita nas métricas de similaridade e agora, usando os metadados. A precisão é calculada através do número de músicas coincidentes do 'ranking' de metadados com os 3 'rankings' de similaridade da *query* dada. Utilizámos a seguinte fórmula para o cálculo da precisão, nesta ainda multiplicamos por 100, para o resultado dar uma percentagem, para ser mais percetível.

$$Precision = \frac{similaridade\ (distâncias) \subset metadados}{total}$$

Neste caso, a **precisão** corresponde a um valor entre 0% e 100%, quanto maior ele for, maior a precisão dos resultados obtidos para os 2 tipos de 'rankings' diferentes gerados. O **total**, da fórmula, corresponde ao número de músicas obtidas no 'ranking', que neste caso, é 10. A **similaridade (distâncias)** é uma matriz obtida que caracteriza o 'ranking' através das distâncias previamente mencionadas. Nos testes efetuados iremos calcular a precisão para as 3 matrizes de distâncias diferentes, a distância Euclidiana, a distância de Manhattan e a distância de Coseno. Por último, os **metadados**, é a matriz que representa o 'ranking' para a *query* usando os metadados como critério de avaliação.

Assim, a precisão obtida para as 3 distâncias foi:

Distância	MT0000414517
Euclidianda	10.0%
Manhattan	20.0%
Coseno	10.0%

Ao analisar esta tabela, conseguimos extrair algumas conclusões, nomeadamente, é notório que para qualquer distância, os valores de precisão são bastantes baixos, sendo o maior cerca de 20.0%, o que nos diz, que apenas 2 músicas que constam dos 'rankings' calculados com as distâncias, também constam no 'ranking' dos metadados, o que traduz numa diferença considerável na forma como os 2 'rankings' (distâncias e metadados) funcionam.

Tal acontece, porque a maneira que as nossas *features* descrevem a música é totalmente distinta da forma como fazem os metadados. As *features* extraem informações de ritmo, compasso, tempo, timbre, variações de energia, potência do sinal, etc., enquanto os metadados são avaliados pelo cantor, quadrante e emoção.

As *features* extraídas caracterizam muito mal as emoções, ao passo que os metadados foram descritos por humanos, que distinguem melhor as emoções. Logo, é completamente natural obtermos uma precisão baixa.

EXERCÍCIO 4

• AVALIAÇÃO SUBJETIVA

O ponto final do projeto consistiu numa avaliação pessoal de cada membro do grupo a cada música presente nos 'rankings' das 3 distâncias, bem como no 'ranking' dos metadados. As avaliações foram feitas conforme a escala de Likert (1 – Muito Má; 2 – Má; 3 – Aceitável; 4 – Boa; 5 – Muito Boa). Fizemos uma avaliação tendo em conta as similaridades das músicas que ouvíamos em relação à *query* correspondente.

Apresentamos, para a *query* dada, a média e o desvio-padrão das avaliações feitas por todos os membros do grupo, assim como a média e o desvio-padrão global de cada 'ranking'. Apresentamos também o valor de precisão de cada 'ranking', que difere da precisão calculada anteriormente, pois consiste na razão entre o número de músicas com avaliação média igual ou acima de 2,5 (recomendação relevante) e o número total de músicas do 'ranking' (10).

• AVALIAÇÃO SUBJETIVA- RANKING EUCLIDIANA

Música	Carolina	Fernanda	Miguel	Rodrigo	Média	Desvio-Padrão
MT0000414517	5	5	5	5	5	0,00
MT0004274911	2	2	1	1	1,5	0,58
MT0003949060	2	3	5	4	3,5	1,29
MT0000218346	2	2	1	1	1,5	0,58
MT0001624303	1	1	2	2	1,5	0,58
MT0003900455	3	3	1	1	2	1,15
MT0004032071	3	3	4	3	3,25	0,50
MT0009208842	2	2	3	2	2,25	0,50
MT0001515531	1	1	1	1	1	0,00
MT0005752234	2	2	2	2	2	0,00
Média	2,3	2,4	2,5	2,2	2,35	
Desvio Padrão	1,16	1,17	1,65	1,40	1,22	
Precisão					0.3(30%)	

• AVALIAÇÃO SUBJETIVA- RANKING MANHATTAN

Música	Carolina	Fernanda	Miguel	Rodrigo	Média	Desvio-Padrão
MT0000414517	5	5	5	5	5	0,00
MT0003949060	2	3	5	4	3,5	1,29
MT0000218346	2	2	1	1	1,5	0,58
MT0004274911	2	2	1	1	1,5	0,58
MT0001624303	3	4	2	2	2,75	0,96
MT0000040632	4	4	3	4	3,75	0,50
MT0034125967	2	2	1	1	1,5	0,58
MT0003900455	3	3	1	1	2	1,15
MT0005469880	2	3	2	2	2,25	0,50
MT0009208842	2	2	3	2	2,25	0,50
Média	2,7	3	2,4	2,3	2,6	
Desvio-Padrão	1,06	1,05	1,58	1,49	1,16	
Precisão					0.4(40%)	

• AVALIAÇÃO SUBJETIVA- RANKING COSENO

Música	Carolina	Fernanda	Miguel	Rodrigo	Média	Desvio-Padrão
MT0000414517	5	5	5	5	5	0,00
MT0004274911	2	2	1	1	1,5	0,58
MT0003949060	2	3	5	4	3,5	1,29
MT0000218346	2	2	1	1	1,5	0,58
MT0001942272	1	1	1	1	1	0,00
MT0002634024	2	3	4	3	3	0,82
MT0009208842	2	2	3	2	2,25	0,50
MT0004032071	3	3	4	3	3,25	0,50
MT0003900455	3	3	1	1	2	1,15
MT0001624303	1	1	2	2	1,5	0,58
Média	2,3	2,5	2,7	2,3	2,45	
Desvio-Padrão	1,16	1,18	1,70	1,42	1,23	
Precisão					0.4(40%)	

- **AVALIAÇÃO SUBJETIVA- RANKING METADADOS**

Música	Carolina	Fernanda	Miguel	Rodrigo	Média	Desvio-Padrão
MT0033397838	2	3	4	4	3,25	0,96
MT0027048677	4	5	5	5	4,75	0,50
MT0000040632	3	3	4	4	3,5	0,58
MT0012331779	4	4	4	5	4,25	0,50
MT0003949060	2	3	3	4	3	0,82
MT0010487769	5	4	4	4	4,25	0,50
MT0010489498	2	3	3	3	2,75	0,50
MT0002222957	4	5	5	5	4,75	0,50
MT0008222676	3	2	3	3	2,75	0,50
MT0007840454	1	1	1	2	1,25	0,50
Média	3	3,3	3,6	3,9	3,45	
Desvio-Padrão	1,25	1,25	1,17	0,99	1,09	
Precisão					0.9(90%)	

- **ANÁLISE DE RESULTADOS**

  - **RANKING EUCLIDIANA**

Ao analisar o 'ranking' percebe se que, em geral, a avaliação não é muito positiva e isto deve se ao facto da existência de grande contraste entre as músicas, em geral, havia músicas que apresentavam um ritmo muito mais lento, instrumentação e arranjos muito diferentes, uso de instrumentos completamente distintos, e que não capturam, nem lembram a essência sonora da música de referência e em destaque está a discrepância entre géneros musicais. Partilhamos da opinião que mais de metade das músicas deste 'ranking' não partilham semelhanças, em especial a faixa MT0001515531, que desenquadra totalmente do resto.

  - **RANKING MANHATTAN**

Passando ao Manhattan, vemos uma evolução em relação á euclidiana, pois o número de músicas com maior discrepância reduziu, vemos ritmos ligeiramente mais parecidos com a query, géneros musicais mais semelhantes e uso de instrumentos com sons mais dentro do mesmo género e som.
Tanto que a maioria das avaliações é superior ou muito perto de 2,5 (metade da escala). É também o ´ranking´ com mais semelhança e melhores recomendações.
Sendo assim o ´ranking de distâncias com melhor precisão tanto na avaliação subjetiva como na objetiva .

- **RANKING COSSENO**

O ‘ranking’ cosseno é muito semelhante ao euclidiana mas mesmo assim é ligeiramente melhor, pois temos mais valores a cima de 2 em relação ao mesmo. Por outro lado neste também vemos diferentes timbres nos instrumentos e as vozes são drasticamente diferentes da referência, também muitas das letras das músicas abordam temas completamente distintos do encontrado na query. Ma,s mesmo assim apresenta uma precisão superior ao ´ranking´ euclidiano, e bastante semelhante ao ´ranking´ manhattan, o que por outro lado é exatamente ao contrário na avaliação objetiva.

- **RANKING METADADOS**

No ´ranking´ metadados é onde há mais coerência entre músicas, tendo a grande maioria um avaliação superior a 3,25. Isto deve se ao facto de na maioria das musicas vermos harmonia e reminiscentes da referência, o tempo e andamento das músicas estão alinhados com as características da referencia o que transmite uma sensação de ritmo e movimento, que são fundamentais para o género musical, que também é bastante semelhante entre as musicas. Por outro lado temos um extremo que é exatamente o contrário, não tem ritmo nenhum, a melodia e harmonia diferem substancialmente da query, a instrumentação é a oposta, o que torna as músicas em dois opostos completos.

• ANÁLISE FINAL

Por fim, se compararmos a avaliação objetiva com a subjetiva, vemos que a precisão é bastante menor na avaliação objetiva. Isto deve se ao facto de estarmos a avaliar apenas 4 opiniões diferentes, que por acaso na grande maioria foram bastante coerentes umas em relação ás outros não havendo discrepâncias significativas entre avaliações. Claro que se a avaliação tivesse maior alcance e os mesmos ´rankings´ fossem avaliados por muitas mais pessoas iria provavelmente descer significativamente.

Em geral , podemos verificar que os resultados da média e do desvio padrão para cada um dos membros do grupo é semelhante. É de realçar que na avaliação subjetiva, os ‘rankings’ das distâncias têm todos, aproximadamente, o mesmo número de precisão (30-40)%, o que nos indica que em 10 músicas, em média 3 a 4 são uma boa recomendação, sendo mais ao menos semelhante em termos de diferenciações de tabelas da avaliação objetiva. Assim, ao comparar as tabelas da distância com a tabela referente aos metadados, verificamos, então, que existem valores de precisão muito mais positivos no ‘ranking’ dos metadados.

CONCLUSÃO

- A extração de *features* com base em características espectrais e temporais dão-nos informações matemáticas/musicais, como, por exemplo: alterações rítmicas, brilho, timbre, dispersão de frequências... não permite captar estruturas harmónicas ou características expressivas, tal como foi feito nos metadados.

- Os metadados foram analisados subjetivamente por humanos, pelo que eles não estão livres de erros.

- Ainda assim, podemos concluir que o uso dos metadados das músicas (artista, género, emoção) foi mais positivo para as recomendações comparativamente às features das mesmas.