



UNIVERSIDADE D
COIMBRA

Music Information Retrieval

Multimédia

Licenciatura em Engenharia Informática
2022/2023

11 de maio de 2023

Autores

João Moreira

✉ joaomoreira@student.dei.uc.pt

👤 2020230563

Rodrigo Figueiredo

✉ rfigueiredo@student.dei.uc.pt

👤 2020236687

Tomás Pinto

✉ tomaspinto@student.dei.uc.pt

👤 2020224069

Índice

1	Introdução	3
2	Extração de features	3
3	Métricas de similaridade	3
3.1	Query 1: MT0000202045.mp3	5
3.2	Query 2: MT0000379144.mp3	6
3.3	Query 3: MT0000414517.mp3	7
3.4	Query 4: MT0000956340.mp3	8
4	Avaliação objetiva	9
4.1	Precisão	9
5	Avaliação subjetiva	11
5.1	Coseno top100	11
5.2	Metadados	12

1. Introdução

Com o avanço da tecnologia e o grande volume de dados disponíveis na *internet*, os *MMIR* tem se tornado cada vez mais relevantes nas nossas vidas. Esta tecnologia 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...

Sendo assim, o objetivo deste trabalho é adquirir a sensibilidade para questões fundamentais dos *MMIR*, tais como: caracterizar e extrair características da multimédia, sobretudo da música, para que seja possível fazer recomendações com base em conteúdo.

2. Extração de features

Tivemos como primeiro objetivo do projeto extrair dados (features) sobre um conjunto de 900 músicas fornecidas no dataset do trabalho.

Estas features foram obtidas de duas maneiras distintas. A primeira consistiu em obter 100 features para cada música através da leitura do ficheiro `top100_features.csv`, onde fizemos a normalização dos valores no intervalo $[0, 1]$ e gravamos em ficheiro a matriz resultante (900×100). A segunda maneira consistiu em utilizar a biblioteca `librosa` para a extração de dados para cada uma das 900 músicas. Os dados obtidos foram `mfcc`, `spectral centroid`, `spectral bandwidth`, `spectral contrast`, `spectral flatness`, `spectral rolloff`, `F0`, `rms`, `zero crossing rate` e `tempo`.

À medida que extraíamos as features referidas, fizemos o cálculo de 7 estatísticas para cada uma. As estatísticas foram a média, desvio padrão, assimetria (*skewness*), curtose (*kurtosis*), mediana, máximo e mínimo.

No final de tudo resultaram 190 features para cada música, e novamente fizemos a normalização destes valores no intervalo $[0, 1]$ e gravamos a matriz em ficheiro (900×190).

As features obtidas por ambos os processos descritos foram importantes no desenvolvimento do projeto.

3. Métricas de similaridade

Nesta etapa do trabalho começamos por desenvolver código para calcular 3 métricas de similaridade, sendo elas a distância Euclidiana, Distância de Manhattan e distância do Coseno.

Para a distância Euclidiana utilizamos o método `linalg.norm()` da biblioteca `numpy`. A fórmula é dada por:

$$D(a, b) = \sqrt{\sum_{i=1}^n (b_i - a_i)^2}$$

Figura 1: Fórmula da distância Euclidiana.

Para a distância de Manhattan utilizamos o método `cityblock()` da biblioteca `scipy`. A formula é dada por:

$$\text{Manhattan Distance } (p, q) = \sum_{i=1}^n |p_i - q_i|$$

Figura 2: Formula da distância de Manhattan.

Por fim, para a distância do Coseno utilizamos o método `cosine()` da biblioteca `scipy`. A formula é dada por:

$$\text{similarity} = \cos(\theta) = \frac{\mathbf{A} \cdot \mathbf{B}}{\|\mathbf{A}\| \|\mathbf{B}\|} = \frac{\sum_{i=1}^n A_i B_i}{\sqrt{\sum_{i=1}^n A_i^2} \sqrt{\sum_{i=1}^n B_i^2}},$$

Figura 3: Formula da distância do Coseno.

Foi nos pedido que criássemos e gravássemos 6 matrizes de similaridade 900x900 para os dois conjuntos de features (top100 e provenientes do librosa) com cada métrica de distância. As matrizes de similaridade indicam a distância entre todas as musicas do dataset.

Com base nas 6 matrizes obtidas, procedemos à geração de 6 rankings de similaridade para 4 queries especificas fornecidas no dataset. Cada ranking vai consistir em 20 sugestões, ou seja, 20 musicas ordenadas com base nos valores das distancias calculados nas matrizes anteriores, sendo a primeira musica a mais semelhante. A propria query em analise não consta nos rankings.

De seguida podemos ver os rankings obtidos para cada query.

3.1. Query 1: MT0000202045.mp3

Euclidiana	Euclidiana	Manhattan	Manhattan	Coseno	Coseno
Top100	Librosa	Top100	Librosa	Top100	Librosa
MT0033841575	MT0005129157	MT0030422114	MT0005129157	MT0033841575	MT0005129157
MT0027002641	MT0012001409	MT0033841575	MT0012001409	MT0027002641	MT0012001409
MT0030487841	MT0002233402	MT0027835071	MT0011899302	MT0030487841	MT0002233402
MT0008575372	MT0011899302	MT0008575372	MT0007799677	MT0008575372	MT0011899302
MT0014576739	MT0007043504	MT0002297016	MT0002233402	MT0014576739	MT0007043504
MT0030422114	MT0007799677	MT0027002641	MT0010624346	MT0030422114	MT0007799677
MT0003390733	MT0004428604	MT0030487841	MT0007043504	MT0003390733	MT0004428604
MT0027835071	MT0011975274	MT0040033011	MT0011975274	MT0027835071	MT0011975274
MT0009188643	MT0002161109	MT0005265641	MT0000732821	MT0011145388	MT0010624346
MT0010617945	MT0010624346	MT0010617945	MT0004428604	MT0009188643	MT0002161109
MT0011145388	MT0007766156	MT0009213083	MT0003787478	MT0010617945	MT0000092267
MT0009213083	MT0009217411	MT0006096934	MT0000092267	MT0009213083	MT0007766156
MT0005331755	MT0000092267	MT0005897799	MT0002161109	MT0005331755	MT0009217411
MT0002233402	MT0026158301	MT0004428604	MT0009217411	MT0002233402	MT0026158301
MT0026727455	MT0003787478	MT0026727455	MT0010344415	MT0026727455	MT0010085729
MT0005265641	MT0005213723	MT0004867185	MT0026158301	MT0005265641	MT0003787478
MT0010344415	MT0001676671	MT0003787478	MT0005213723	MT0010344415	MT0005213723
MT0000711493	MT0027002641	MT0001340713	MT0005039941	MT0000711493	MT0001676671
MT0004428604	MT0010344415	MT0009188643	MT0002379222	MT0018029465	MT0027002641
MT0018029465	MT0001942272	MT0011145388	MT0000315392	MT0005469880	MT0010344415

Tabela 1: Rankings de similaridade para a query 1

3.2. Query 2: MT0000379144.mp3

Euclidiana	Euclidiana	Manhattan	Manhattan	Coseno	Coseno
Top100	Librosa	Top100	Librosa	Top100	Librosa
MT0009010830	MT0006540794	MT0015005100	MT0004141823	MT0009010830	MT0006540794
MT0003778826	MT0004141823	MT0009010830	MT0011957429	MT0003778826	MT0004141823
MT0015005100	MT0011957429	MT0031996897	MT0008222676	MT0015005100	MT0011957429
MT0028627699	MT0008222676	MT0012124855	MT0006540794	MT0028627699	MT0008222676
MT0012124855	MT0009845271	MT0028627699	MT0009845271	MT0012124855	MT0002222957
MT0000044741	MT0012041920	MT0003778826	MT0002222957	MT0027035970	MT0009845271
MT0005478759	MT0002222957	MT0000992846	MT0011821215	MT0000044741	MT0006510599
MT0027035970	MT0006510599	MT0001376988	MT0000044741	MT0003794106	MT0003794106
MT0003794106	MT0003778826	MT0000044741	MT0002119439	MT0005478759	MT0012041920
MT0000992846	MT0011821215	MT0005478759	MT0003262589	MT0002262181	MT0005515169
MT0031996897	MT0000044741	MT0027035970	MT0012041920	MT0031996897	MT0007379559
MT0001376988	MT0007379559	MT0003794106	MT0003778826	MT0000992846	MT0011999773
MT0009208842	MT0006164654	MT0005737276	MT0006164654	MT0001376988	MT0002479795
MT0002262181	MT0003669917	MT0003903675	MT0010617945	MT0009208842	MT0009041370
MT0005737276	MT0008684922	MT0002262181	MT0006510599	MT0010465830	MT0003778826
MT0010465830	MT0005202791	MT0009208842	MT0005202791	MT0005737276	MT0003669917
MT0018031959	MT0003262589	MT0008222676	MT0003669917	MT0018031959	MT0005202791
MT0033958450	MT0005515169	MT0006367176	MT0005523766	MT0033958450	MT0003462877
MT0012041920	MT0003462877	MT0006540794	MT0003794106	MT0008222676	MT0006367176
MT0012396528	MT0009041370	MT0015742096	MT0010465830	MT0012041920	MT0003262589

Tabela 2: Rankings de similaridade para a query 2

3.3. Query 3: MT0000414517.mp3

Euclidiana	Euclidiana	Manhattan	Manhattan	Coseno	Coseno
Top100	Librosa	Top100	Librosa	Top100	Librosa
MT0003900455	MT0003949060	MT0003900455	MT0003949060	MT0003900455	MT0004274911
MT0000203193	MT0004274911	MT0003949060	MT0004274911	MT0009897495	MT0003949060
MT0009897495	MT0001515531	MT0000203193	MT0000040632	MT0000203193	MT0001515531
MT0009521580	MT0003900455	MT0009521580	MT0000218346	MT0000901959	MT0002634024
MT0018651126	MT0009897495	MT0013955066	MT0003900455	MT0009521580	MT0003900455
MT0013955066	MT0000040632	MT0000218346	MT0005469880	MT0013955066	MT0009897495
MT0034186195	MT0004032071	MT0001333258	MT0008401073	MT0000888329	MT0001942272
MT0000901959	MT0005469880	MT0009897495	MT0001515531	MT0004428604	MT0004032071
MT0004428604	MT0007043936	MT0034186195	MT0001624303	MT0018651126	MT0000040632
MT0009346128	MT0034005433	MT0001703346	MT0034125967	MT0009346128	MT0005752234
MT0000888329	MT0005752234	MT0018651126	MT0009897495	MT0002379222	MT0005469880
MT0000218346	MT0009208842	MT0034005433	MT0005752234	MT0034186195	MT0009208842
MT0012331779	MT0000636335	MT0013633209	MT0004032071	MT0017667847	MT0007043936
MT0034005433	MT0002634024	MT0005469880	MT0009208842	MT0000218346	MT0009346128
MT0013633209	MT0001624303	MT0017797643	MT0001942272	MT0000040632	MT0034005433
MT0002379222	MT0008401073	MT0034125967	MT0001236649	MT0002634024	MT0000636335
MT0001703346	MT0001942272	MT0009346128	MT0001109401	MT0012331779	MT0001624303
MT0005752234	MT0034125967	MT0013161246	MT0030282917	MT0001703346	MT0002698706
MT0017667847	MT0009346128	MT0005752234	MT0006769480	MT0027835071	MT0001109401
MT0009991160	MT0030282917	MT0000040632	MT0034005433	MT0009991160	MT0008401073

Tabela 3: Rankings de similaridade para a query 3

3.4. Query 4: MT0000956340.mp3

Euclidiana	Euclidiana	Manhattan	Manhattan	Coseno	Coseno
Top100	Librosa	Top100	Librosa	Top100	Librosa
MT0003106472	MT0006096934	MT0003106472	MT0006096934	MT0003106472	MT0006096934
MT0010736208	MT0003787478	MT0010736208	MT0005625762	MT0010736208	MT0003787478
MT0014703649	MT0005625762	MT0014703649	MT0003787478	MT0014703649	MT0003603772
MT0005409948	MT0040033011	MT0014615863	MT0040033011	MT0005409948	MT0005625762
MT0014615863	MT0003603772	MT0005409948	MT0033177286	MT0014615863	MT0040033011
MT0013612461	MT0033177286	MT0004028719	MT0003603772	MT0004882280	MT0033177286
MT0004032071	MT0004867185	MT0002372242	MT0006640142	MT0013612461	MT0003213835
MT0010615428	MT0013389935	MT0007535042	MT0008511909	MT0003724610	MT0008511909
MT0003724610	MT0008511909	MT0004032071	MT0013389935	MT0002372242	MT0004867185
MT0040033011	MT0005331755	MT0004085907	MT0005331755	MT0004032071	MT0013389935
MT0002372242	MT0003213835	MT0003724610	MT0003213835	MT0010615428	MT0011145388
MT0004882280	MT0006640142	MT0004942017	MT0001217651	MT0040033011	MT0004958762
MT0009800907	MT0004958762	MT0013612461	MT0004867185	MT0007535042	MT0005331755
MT0004293364	MT0011145388	MT0010615428	MT0004942017	MT0004293364	MT0006640142
MT0004028719	MT0001217651	MT0040033011	MT0001680969	MT0009800907	MT0014576739
MT0007535042	MT0001680969	MT0035334027	MT0005265641	MT0004028719	MT0008570712
MT0030036616	MT0004942017	MT0011697297	MT0005409948	MT0030036616	MT0007535042
MT0014794891	MT0005265641	MT0004882280	MT000857071	MT0004085907	MT0001217651
MT0013822237	MT0014576739	MT0009800907	MT0011145388	MT0010624346	MT0006367176
MT0010624346	MT0030422114	MT0010624346	MT0014576739	MT0014794891	MT0005265641

Tabela 4: Rankings de similaridade para a query 4

Analisando as tabelas obtidas podemos perceber uma semelhança entre os rankings de similaridade das 3 distâncias para as top100 features de cada query, verificando que os rankings são quase inteiramente compostos pelas mesmas musicas apenas variando a posição em que as mesmas são apresentadas. A mesma conclusão pode ser tirada para para os rankins das 3 distâncias relativas às features do librosa.

Comparando os rankings das top100 features com os rankings das features do librosa em cada distância, estes mostram-se completamente diferentes sendo raro encontrar as mesmas musicas nos dois rankings.

Através destas observações podemos concluir que as diferenças de features tiveram um papel mais impactante na construção dos rankings de similaridade comparativamente as diferenças de distancias que resultaram em poucas mudanças nos mesmos. Assim concluímos que é mais importante procurar boas features descritivas do que métricas de similaridade.

4. Avaliação objetiva

Chegando à parte de avaliação, obtemos novos rankings de recomendação para cada query mas desta vez baseados em metadados específicos, sendo estes o artista, género, quadrante e emoção.

Extraímos os metadados disponibilizados no dataset e procedemos ao cálculo de similaridades entre todas as músicas da base de dados, originando uma matriz 900x900, onde o valor de similaridade para cada duas músicas consiste nos itens coincidentes entre ambos, ou seja, se tanto a música de referência como o alvo tiverem género = jazz e tiverem como autor o mesmo artista o valor será 2.

Tendo em conta a matriz de similaridade anterior, geramos para cada query um ranking das 20 músicas com maior número de similaridade com a mesma.

Os resultados foram os seguintes:

Musica	MT0000202045	MT0000379144	MT0000414517	MT0000956340
1	MT0012862507	MT0031951901	MT0010489498	MT0016743722
2	MT0014475915	MT0014584473	MT0010487769	MT0013885218
3	MT0000888329	MT0013416300	MT0027048677	MT0014615863
4	MT0003022328	MT0013080259	MT0000040632	MT0004293364
5	MT0001494812	MT0011032905	MT0012331779	MT0033097471
6	MT0031898123	MT0005157391	MT0033397838	MT0014845647
7	MT0007556029	MT0005253065	MT0003949060	MT0001927746
8	MT0004867564	MT0008170600	MT0002222957	MT0004459450
9	MT0011922905	MT0007652281	MT0010900969	MT0012914763
10	MT0030369896	MT0007349999	MT0007840454	MT0012001409
11	MT0007453719	MT0007338724	MT0008222676	MT0010897525
12	MT0034186620	MT0001526386	MT0009202768	MT0009217411
13	MT0001058887	MT0004287283	MT0001703346	MT0001891229
14	MT0003025046	MT0005115042	MT0009010830	MT0029099688
15	MT0004850690	MT0004131058	MT0005713768	MT0014817509
16	MT0005285696	MT0001929641	MT0002053300	MT0032957418
17	MT0034577404	MT0001934726	MT0002033629	MT0005897799
18	MT0011938737	MT0003863509	MT0015005100	MT0015962332
19	MT0002846256	MT0029877658	MT0008684922	MT0015541501
20	MT0011051663	MT0003114552	MT0004669603	MT0002372242

Tabela 5: Ranking dos metadados para as 4 queries.

4.1. Precisão

Procedemos ao cálculo da precisão dos 4 rankings apresentados na tabela 5. A precisão é calculada através do número de músicas coincidentes do ranking dos metadados com os 24 rankings baseados nas distâncias calculadas anteriormente (tabelas 1-4), sobre o total de músicas dos rankings (20).

Os resultados obtidos foram os seguintes:

	MT0000202045	MT0000379144	MT0000414517	MT0000956340
Euclidiana top100	0	0	0.1	0.15
Manhattan top100	0	0	0.15	0.1
Cosseno top100	0	0	0.15	0.15
Euclidiana librosa	0	0	0.1	0
Manhattan librosa	0	0	0.1	0
Cosseno librosa	0	0	0.1	0

Tabela 6: Precisão dos ranking dos metadados relativamente aos rankings calculados na pergunta 3 por queries

Analisando os dados da tabela 6, podemos reparar que para as duas primeiras queries não houve qualquer correlação dos rankings relativos aos metadados com os rankings das distâncias. As duas queries restantes já apresentaram alguma correlação, ainda que muito baixa, sendo o maior valor de precisão registado de 0.15 equivalente a 3 músicas coincidentes.

No geral podemos verificar valores de precisão muito baixos (ou mesmo 0), o que nos permite perceber as grandes diferenças entre a avaliação de similaridade recorrendo à avaliação dos metadados ou recorrendo as features das músicas, diferenças essas incidentes nos tipos de características avaliadas em cada abordagem, que são totalmente diferentes.

Tal acontece, porque a forma como 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... enquanto que 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, cujos distinguem melhor as emoções. Logo, é completamente natural obtermos uma precisão baixa.

5. Avaliação subjetiva

O ponto final do projeto consistiu numa avaliação pessoal de cada membro do grupo a cada música presente nos rankings do Coseno top100 bem como nos rankings dos metadados relativos as 4 queries. As avaliações foram feitas de acordo com 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 musicas que ouvíamos em relação à query correspondente.

Apresentamos, para cada query, 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 diferentemente da precisão calculada na tabela 6, 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 (20).

5.1. Coseno top100

	MT0000202045		MT0000379144		MT0000414517		MT0000956340	
	média	desvio	média	desvio	média	desvio	média	desvio
João	1.75	1.118	2.6	1.569	1.7	1.341	3.35	1.598
Rodrigo	1.8	1.240	2.75	1.409	1.5	1.100	3.45	1.605
Tomás	2.05	1.146	3.05	1.234	1.75	0.850	3.55	1.571

Tabela 7: Média e desvio de todos os membros por query para o ranking coseno top100

	MT0000202045	MT0000379144	MT0000414517	MT0000956340
Média	1.866667	2.8	1.65	3.45
Desvio	1.156656	1.399758	1.102001	1.566871
Precisão	0.15 (15%)	0.55 (55%)	0.15 (15%)	0.75 (75%)

Tabela 8: Média, desvio padrão e precisão para o ranking coseno top100

Pela análise dos resultados da tabela 7 podemos verificar que os resultados da média e desvio padrão de cada query para cada um dos membros do grupo estão muito parecidos, o que indica concordância e semelhança nas opiniões relativamente a se as musicas são ou não boas recomendações.

Analisando a tabela 8 percebemos que a quarta query apresenta a maior média, e consequentemente a maior precisão de entre as 4 queries (75%), indicando que, aos nossos olhos, a lista de recomendações desta query é a melhor. Por outro lado a primeira e a terceira queries apresentaram valores muito baixos de precisão (15%), possuindo assim para nos, listas de recomendação más, com poucas musicas realmente parecidas à query em questão. A segunda query apresentou valores medianos de precisão (50%), e assim várias músicas parecidas e várias sem relação alguma.

5.2. Metadados

	MT0000202045		MT0000379144		MT0000414517		MT0000956340	
	média	desvio	média	desvio	média	desvio	média	desvio
João	1.65	0.988	4.15	0.670	1.609	2.8	2.55	1.276
Rodrigo	1.5	1.000	4.35	0.812	1.719	2.7	2.6	1.535
Tomás	1.7	1.174	3.6	1.046	1.589	3	3.05	1.700

Tabela 9: Média e desvio de todos os membros por query para o ranking cosseno top100

	MT0000202045	MT0000379144	MT0000414517	MT0000956340
Média	1.616667	4.033333	2.833333	2.733333
Desvio	1.043002	0.901349	1.617349	1.505545
Precisão	0.15 (15%)	0.95 (95%)	0.5 (50%)	0.55 (55%)

Tabela 10: Média, desvio padrão e precisão para o ranking de metadados

Pela análise dos resultados da tabela 9 verificamos que, à semelhança dos valores da tabela 7 relativos ao Cosseno top100, resultados da média e desvio padrão de cada query para cada um dos membros do grupo estão muito parecidos, mais uma vez, indicando semelhança nas opiniões dos elementos do grupo.

Analisando a tabela 10 percebemos que agora a segunda query apresenta a maior precisão de entre as 4 queries (95%), indicando que praticamente todas as músicas são recomendações relevantes. A terceira e quarta queries apresentaram valores medianos de precisão (50% e 55% respetivamente). A primeira query apresentou valores maus de precisão (15%), e assim uma má lista de recomendações.

Comparando os resultados do Cosseno top100 com os metadados (tabelas 8 e 10), verificamos no conjunto, valores de precisão mais positivos nos rankings dos metadados, onde a precisão relativa à segunda e terceiras queries sobe, mesmo verificando uma descida na precisão da quarta query.

Ao ouvir as recomendações da quarta *query* reparamos que as músicas são maioritariamente do género: *RAP* e apresentam ritmicidade, volume, intensidade e timbres semelhantes. O resultado de 75% de precisão do Cosseno top100 deve-se ao facto de as nossas *features* serem boas para definir tais características.

Existe algumas razões para tal acontecer:

- 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.

O facto do dataset possuir um numero relativamente pequeno de musicas pode contribuir para o facto dos valores de precisão não serem todos muito altos, como por exemplo a primeira query, que mostrou valores de precisão maus em ambos os resultados.

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