



FACULDADE DE
CIÊNCIAS E TECNOLOGIA
UNIVERSIDADE DE
COIMBRA

Music Information Retrieval

TRABALHO PRÁTICO 2

Ricardo Guegan, nº 202021158
Samuel Machado, nº 2020

Multimédia

Ano letivo 2022/23

Introdução

Este projeto teve como objetivo perceber o funcionamento de um sistema de recomendação musical. As duas principais componentes deste sistema de recomendação são, os conteúdos das músicas conseguido através da análise dos sinais de áudio e a análise de Metadados, que se focam principalmente nas emoções presentes em cada música.

Exercício 1

Na primeira semana do projeto analisamos e pesquisamos sobre sistemas de recomendação real, mais especificamente o funcionamento do sistema de recomendação do Spotify. Sobre este aprendemos que a sua recomendação é baseada em 3 grandes etapas, os dados que recolhem do utilizador, como o seu histórico de músicas e as suas favoritas, com estes dados vai ser feito:

1- “Collaborative Filtering”, ou seja, com base nos outros utilizadores da aplicação vão tentar encontrar perfis semelhantes ao nosso e recomendar as músicas que estes ouvem. (pesquisa com Metadados)

2- Análises do Áudio estas visam e analisar as features das músicas ouvidas e procurar outras com features semelhantes. (pesquisa conteúdo)

Exercício 2

No exercício 2, o objetivo é analisar os sinais de áudio das músicas e a partir destas extrair as features. Para a extração das features usamos a biblioteca do Librosa.

As features que vamos contém informação sobre o seguinte:

- Features para a análise do Timbre: Mel-Frequency Cepstral Coefficients, spectral centroid, spectral bandwidth, spectral contrast, spectral flatness e spectral rolloff
- Features para a análise da intensidade: Root Mean Square
- Features para a análise da frequência: fundamental frequency, zero-crossing rate
- Features para a análise do ritmo: Tempo

Como estas features são aplicadas a janelas de 2048 samples (valor default) obtemos um grande conjunto de valores com os quais é difícil operar a nível computacional e fazer comparações com outras músicas pois são demasiados específicos. Para conseguir então manipular melhor estas features temos de as conseguir representar em apenas alguns números, para tal vamos recorrer às seguintes estatísticas. Média, desvio padrão, assimetria, curtose, mediana, máximo, mínimo.

Com este processo, estamos a descrever uma música em um conjunto de 190 valores. Para facilitar também a comparação com outras músicas, estes 190 valores são normalizados em um range de [0,1] o que vai fazer com que no momento do cálculo das distâncias todos os valores tenham o mesmo peso.

Exercício 3

No exercício 3, vamos fazer os rankings de similaridades com base no conteúdo das músicas, para o fazer vamos partir do pressuposto que todas as features têm o mesmo peso. Para fazer a comparação entre as músicas vamos usar 3 medidas de comparação: a distância Euclidiana, a distância de Manhattan e a distância dos cossenos.

Com as nossas Features:

Euclidean distance matches:

['MT0000202045.mp3', 'MT0005129157.mp3', 'MT0012001409.mp3', 'MT0002233402.mp3', 'MT0011975274.mp3', 'MT0010624346.mp3', 'MT0002161109.mp3', 'MT0011899302.mp3', 'MT0007043504.mp3', 'MT0000092267.mp3', 'MT0026158301.mp3', 'MT0009217411.mp3', 'MT0003787478.mp3', 'MT0008979040.mp3', 'MT0011376343.mp3', 'MT0002379222.mp3', 'MT0000315392.mp3', 'MT0007535042.mp3', 'MT0007799677.mp3', 'MT0007766156.mp3', 'MT0002811975.mp3']

Manhattan distance matches:

['MT0000202045.mp3', 'MT0005129157.mp3', 'MT0012001409.mp3', 'MT0010624346.mp3', 'MT0011975274.mp3', 'MT0002233402.mp3', 'MT0011899302.mp3', 'MT0000092267.mp3', 'MT0007043504.mp3', 'MT0007799677.mp3', 'MT0002161109.mp3', 'MT0003787478.mp3', 'MT0000732821.mp3', 'MT0002379222.mp3', 'MT0009217411.mp3', 'MT0026158301.mp3', 'MT0004428604.mp3', 'MT0000315392.mp3', 'MT0011376343.mp3', 'MT0007535042.mp3', 'MT0005039941.mp3']

Cosine similarity matches:

['MT0000202045.mp3', 'MT0012001409.mp3', 'MT0005129157.mp3', 'MT0002233402.mp3', 'MT0010624346.mp3', 'MT0011975274.mp3', 'MT0000092267.mp3', 'MT0002161109.mp3', 'MT0011899302.mp3', 'MT0007043504.mp3', 'MT0026158301.mp3', 'MT0007535042.mp3', 'MT0003787478.mp3', 'MT0009217411.mp3', 'MT0008979040.mp3', 'MT0011376343.mp3', 'MT0002379222.mp3', 'MT0000315392.mp3', 'MT0007766156.mp3', 'MT0007799677.mp3', 'MT0002811975.mp3']

FeaturesTop100:

Euclidean distance matches:

['MT0000202045.mp3', 'MT0033841575.mp3', 'MT0027002641.mp3', 'MT0030487841.mp3', 'MT0008575372.mp3', 'MT0014576739.mp3', 'MT0030422114.mp3', 'MT0003390733.mp3', 'MT0027835071.mp3', 'MT0009188643.mp3', 'MT0010617945.mp3', 'MT0011145388.mp3', 'MT0009213083.mp3', 'MT0005331755.mp3', 'MT0002233402.mp3', 'MT0026727455.mp3', 'MT0005265641.mp3', 'MT0010344415.mp3', 'MT0000711493.mp3', 'MT0004428604.mp3', 'MT0018029465.mp3']

Manhattan distance matches:

['MT0000202045.mp3', 'MT0030422114.mp3', 'MT0033841575.mp3', 'MT0027835071.mp3', 'MT0008575372.mp3', 'MT0002297016.mp3', 'MT0027002641.mp3', 'MT0030487841.mp3', 'MT0040033011.mp3', 'MT0005265641.mp3', 'MT0010617945.mp3', 'MT0009213083.mp3', 'MT0006096934.mp3', 'MT0005897799.mp3', 'MT0004428604.mp3', 'MT0026727455.mp3', 'MT0004867185.mp3', 'MT0003787478.mp3', 'MT0001340713.mp3', 'MT0009188643.mp3', 'MT0011145388.mp3']

Cosine similarity matches:

['MT0000202045.mp3', 'MT0033841575.mp3', 'MT0027002641.mp3', 'MT0030487841.mp3', 'MT0008575372.mp3', 'MT0014576739.mp3', 'MT0030422114.mp3', 'MT0003390733.mp3', 'MT0027835071.mp3', 'MT0011145388.mp3', 'MT0009188643.mp3', 'MT0010617945.mp3', 'MT0009213083.mp3', 'MT0005331755.mp3', 'MT0002233402.mp3', 'MT0026727455.mp3', 'MT0005265641.mp3', 'MT0010344415.mp3', 'MT0000711493.mp3', 'MT0018029465.mp3', 'MT0005469880.mp3']

Como podemos ver, para a música MT0000202045.mp3, as diferentes distâncias devolvem resultados bastante semelhantes, tendo quase que o mesmo top 10 mas numa ordem diferente concluímos com isto que apesar da escolha da distância ter a sua importância aquilo que tem maior peso é a escolha das features pois entre as nossas features e as features top100 na qual o top 10 é totalmente diferente e no top 20 apenas tem 2 músicas em comum.

Exercício 4.1

Neste exercício assumimos fazemos recomendação baseada em metadados, isto é feito com um sistema de pontos, no qual para os seguintes dados (Artista, Quadrante, emoção/*mood*, género) por cada um que as músicas têm em comum com a música query esta recebe um ponto.

Distâncias:	Euclidiana	Manhattan	Cos	Euclidiana	Manhattan	Cos
MT0000202045.mp3	0	0	0	0	0	0
MT0000379144.mp3	0	0	0	0	0	0
MT0000956340.mp3	0.1	0.1	0.1	0.1	0.15	0.15
MT0000414517.mp3	0	0	0	0.15	0.1	0.15

Precisão das diferentes distâncias (Nossas Features e Top100)

Ao analisar os valores de *precision* nos quais confrontamos os rankings de similaridade obtidos com a pesquisa por conteúdo com os rankings de similaridade obtidos com os metadados, vemos que estes diferem imenso, em alguns cenários não tendo sequer um em comum. Isto é expectável, pois os metadados utilizados baseiam-se em emoção e isso é algo difícil de extrair com features e também porque as features utilizadas são de baixo nível.

Exercício 4.2

Neste exercício, vamos ser nós a avaliar com base na escala de Likert os top 20 dos resultados do Top100 com as distâncias dos Cossenos e com os Metadados.

Fig. 1 - Média das avaliações subjetivas das Features Top100

1	5.0	2.0	2.6	4.3	1.6	1.3	1.3	2.0	1.0	1.3	2.0	2.0	2.3	3.0	1.6	2.0	1.0	2.6	1.3	1.3	2.0
2	5.0	2.6	3.6	2.0	3.3	3.0	2.6	2.3	4.6	3.3	2.6	2.6	3.6	3.6	3.6	3.3	4.3	3.3	2.0	4.0	3.0
3	5.0	4.3	5.0	5.0	4.3	4.3	2.3	2.6	2.0	1.6	4.6	1.6	3.6	2.6	2.0	2.0	3.6	4.3	3.6	1.6	1.6
4	5.0	4.3	5.0	5.0	4.3	4.3	2.3	2.6	2.0	1.6	4.6	1.6	3.6	2.6	2.0	2.0	3.6	4.3	3.6	1.6	1.6

Fig. 2 - Média das avaliações subjetivas dos Metadados

1	5.0	2.3	2.6	2.6	3.3	1.6	2.3	2.6	4.3	1.3	3.3	2.3	1.6	1.0	1.6	2.3	3.3	1.3	2.6	1.3	2.3
2	5.0	2.6	2.6	3.0	3.6	2.3	4.6	3.6	1.6	1.6	2.0	2.0	2.6	2.3	2.3	3.3	4.0	1.0	3.6	3.0	3.6
3	5.0	3.3	2.6	1.6	1.3	2.0	3.0	3.0	2.6	3.0	3.0	2.0	2.6	1.3	1.3	3.0	2.0	3.3	3.3	2.6	2.6
4	5.0	4.6	4.3	4.0	4.0	4.0	4.0	2.6	2.3	3.3	3.3	1.0	2.3	3.0	1.6	3.6	2.0	2.3	3.3	2.6	3.3

Legenda:

1. MT0000202045.mp3
2. MT0000379144.mp3
3. MT0000956340.mp3
4. MT0000414517.mp3

Features Top100:

Músicas	Médias	Desvios Padrão	Precision
MT0000202045.mp3	2.46031746	1.13843943	0.45
MT0000379144.mp3	2.9047619	1.24357761	0.65
MT0000956340.mp3	2.58730159	1.10747705	0.7
MT0000414517.mp3	3.07936508	1.25105146	0.7
Total	2.75793651	1.21209861	

Metadados:

Músicas	Médias	Desvios Padrão	Precision
MT0000202045.mp3	2.11111111	1.15622691	0.25
MT0000379144.mp3	3.28571429	1.18761561	0.9
MT0000956340.mp3	2.92063492	1.30081278	0.55
MT0000414517.mp3	3.25396825	1.36821346	0.65
Total	2.89285714	1.34239058	

Como podemos ver pelos resultados acima, aqueles que obtiveram resultados com mais precisão foram os que utilizaram as features Top100, pois estas fazem alusão a algo menos subjetivo. Pela mesma razão, o desvio padrão nesta situação é também menor, apesar de as médias terem sido mais altas nos resultados dos metadados. O sample que estamos a analisar é pequeno, mas podemos concluir que os níveis de *precision* para as músicas são bastante positivos e que nas features, músicas como a query 4 na qual o timbre é mais distinto consegue obter uma melhor pontuação utilizando as features, pois várias delas servem para analisar o timbre. O que é importante notar, é que em ambas as avaliações, a concentração de resultados positivos encontra-se no top 10, o que nos permite concluir que a utilização de ambas as técnicas de recomendação, por metadados e por conteúdos, devem ser usadas em conjunto e que os níveis de *precision* seriam bem mais altos se juntássemos os top 10 de cada música.