

Primeiro exercício de programação

Recomendações Musicais

Dezembro/2020

I. INTRODUÇÃO

Sistemas de Recomendação Musical (SRM) são ferramentas que recomendam músicas, artistas ou gêneros para um usuário específico a partir de determinadas regras baseadas no interesse pessoal do usuário. Para Resnick and Varian [1997], sistemas de recomendações são um dos principais fatores que as plataformas de *Streaming* de músicas se diferenciam das mídias tradicionais como, como o rádio, já que a possibilidade de um conteúdo construído exclusivamente para cada usuário torna o *streaming* mais interessante. Sistemas de Recomendação formam uma área muito estudada no campo de aprendizado de máquina e, embora existam diversas técnicas, uma etapa comum a qualquer estudo é a caracterização inicial dos dados. Ter um panorama global dos dados ajuda a ter *insights* de novas ideias baseadas no comportamento desses dados, que pode ajudar, por exemplo, a direcionar publicidade e eventos.

II. OBJETIVOS

Neste exercício, você deverá analisar os dados do Spotify¹, um dos serviços de *streaming* de música mais popular e usado do mundo². Você buscará criar um panorama da amostra disponibilizada e tentará responder as seguintes perguntas:

1. Existe uma mudança de preferência na duração das músicas com o passar dos anos? Apresente a relação da duração mínima, média e máxima das músicas em função dos anos dessa amostra. Considere o último *timestamp* de alteração da lista de reprodução para responder essa pergunta.
2. O que é mais comum, *playlists* onde existem mais músicas distintas de um mesmo artista (considere mais de 50%) ou *playlists* mais diversificadas?
3. Para o artista mais popular (isto é, presente no maior número de *listas*), crie um gráfico que relacione o número de *playlists* distintas em que esse artista apareceu com o passar dos anos.
4. Usando as 50 *playlists* com o maior número de seguidores, crie um sistema simples de recomendação capaz de indicar uma única *playlist* com base em um conjunto de músicas passadas por parâmetro. Para isso, utilize o arquivo de entrada *recomendacao-playlist*, no Lemonade, contendo um conjunto de músicas para o qual você deverá indicar uma lista.

¹<http://www.spotify.com>

²https://en.wikipedia.org/wiki/Comparison_of_on-demand_music_streaming_services

III. BASE DE DADOS

A base de dados utilizada representa uma extrao das listas das msicas escutadas no Spotify, obtida pelo *The Million Song Dataset* [Bertin-Mahieux et al., 2011]. A base original contm 1.000.000 de *playlists* criadas por usurios na plataforma Spotify e foi coletada por pesquisadores interessados em explorar como melhorar a experincia de escuta musical. Essas listas de reproduo foram criadas durante o perodo de janeiro de 2010 a outubro de 2017. Cada lista de reproduo contm um ttulo da lista de reproduo, a lista de faixas (incluindo metadados da faixa), informaes de edio (ltima edio, nmero de edies da lista de reproduo) e outras informaes diversas sobre o lista de reproduo.

A base de dados disponibilizada para esse trabalho prtico, e j configurada no Lemonade, contm uma amostra de 191.000 listas e um total de 967.554 msicas distintas. A base no Lemonade foi pr-processada, e se encontra na forma de dois arquivos, o `spotify_playlist` lista as informaes da *playlist* e o arquivo `spotify_tracks` lista as informaes da msicas. Especificamente, o arquivo `spotify_playlist` encontra-se em formato *csv* com as colunas:

- *pid* - o id da lista de reproduo;
- *name* - o nome da lista de reproduo;
- *modified_at* - *timestamp* (em segundos desde a poca) quando esta lista de reproduo foi atualizada pela ltima vez;
- *duration_ms* - a durao total de todas as faixas da lista de reproduo (em milissegundos);
- *num_albums* - o nmero de lbuns exclusivos para as faixas na lista de reproduo;
- *num_followers* - o nmero de seguidores que esta lista de reproduo tinha no momento da criao da base;
- *num_artists* - o nmero total de artistas nicos para as faixas da lista de reproduo;
- *num_edits* - o nmero de sesses de edio distintas;
- *num_tracks* - o nmero de faixas na lista de reproduo.

J o arquivo `spotify_tracks`, tambm em *csv*, apresenta os seguintes campos:

- *pid_playlist* - o id da lista de reproduo  qual pertence a faixa;
- *track_name* - o nome da faixa;
- *track_uri* - o URI Spotify da faixa;
- *album_name* - o nome do lbum da faixa;
- *album_uri* - o URI Spotify do lbum;
- *artist_name* - o nome do artista principal da faixa;
- *artist_uri* - o URI do Spotify do artista principal da faixa;
- *duration_ms* - a durao da faixa em milissegundos;
- *pos* - a posio da faixa na lista de reproduo (base zero);

IV. DOCUMENTA  O E PAR METROS DE AVALIA  O

Dever  ser escrito um relat rio para explicar as an lises, os resultados produzidos e como cada an lise foi constru da no Lemonade. O aluno dever  entregar um  nico arquivo compactado contendo o relat rio produzido com as imagens dos fluxos criados, junto com seus arquivos *json* produzidos pelo Lemonade. Ser  criada uma tarefa no moodle para a entrega.

Foi criado um f rum espec fico no moodle para discuss o de d vidas sobre o exerc cio. Publique suas d vidas l , assim as respostas j  v o ajudar a todos. Sendo assim, n o esque a de conferir as respostas j  publicadas para ver se a resposta para sua d vida j  n o est  l .

Aproveite essa oportunidade para conhecer um pouco do Spark observando o c digo-fonte produzido durante a execu  o (n o se preocupe, vamos falar sobre Spark em janeiro). Al m disso, ao elaborar essa atividade, voc  encontrar  muitos conceitos comuns na  rea de *Data Science/Big Data Analytics*. Caso tenha alguma sugest o para a plataforma do Lemonade, fique   vontade para incluir no relat rio.

REFER NCIAS

Thierry Bertin-Mahieux, Daniel P.W. Ellis, Brian Whitman, and Paul Lamere. The million song dataset. In *Proceedings of the 12th International Conference on Music Information Retrieval (ISMIR 2011)*, 2011.

Paul Resnick and Hal R. Varian. Recommender systems. *Commun. ACM*, 40(3):56–58, March 1997. ISSN 0001-0782. doi: 10.1145/245108.245121. URL <https://doi.org/10.1145/245108.245121>.

M os   obra (ah, e um Feliz Natal e um  timo ano novo)!

“When life gives you lemons, make lemonade”

(Proverbial phrase)