

Proposta Projeto Sistemas de Recomendação 2019/1: Predições de Músicas para Novas *Playlists* baseadas em Filtragem Colaborativa de Itens

Breno Martins
Universidade Federal de Minas Gerais
Belo Horizonte, Minas Gerais
brenoam@dcc.ufmg.br

Marina Moreira
Universidade Federal de Minas Gerais
Belo Horizonte, Minas Gerais
marinamoreira@dcc.ufmg.br

Ronald Pereira
Universidade Federal de Minas Gerais
Belo Horizonte, Minas Gerais
ronald.pereira@dcc.ufmg.br

1 PROPOSTA

O nosso grupo optou pela opção de “novelty track”, desenvolvendo novas estratégias de análises e engenharia de dados para o código disponível publicamente¹. Esse sistema de recomendação foi originalmente desenvolvido em função de uma proposta de projeto² pela empresa de *streaming* de músicas Spotify no ano de 2018 na conferência da ACM (*Association for Computing Machinery*).

O interesse do grupo por esse assunto se deu em parte pelo interesse numa análise, aprendizado baseado em aplicação e possível melhoria de algoritmos com objetivo de se recomendar músicas similares para um determinado usuário que realiza a utilização de uma plataforma de *streaming* de músicas, de modo a auxiliá-lo a criar uma *playlist* mais robusta, baseada em recomendações realizadas. A outra parte complementar desse interesse se deu pelo fato de ter sido um desafio recente lançado por uma das maiores empresas atualmente do mercado de *streaming* de música, tornando mais interessante a análise dos possíveis resultados alcançados. Além disso, esse assunto se encaixa justamente no conteúdo de sistemas de recomendação baseados em filtragem colaborativa de itens, que foi o conteúdo lecionado na aula do dia 02 de Abril de 2019 pelo professor Rodrygo, se mostrando uma excelente oportunidade de se aplicar um conhecimento.

A nossa proposta será de desenvolver uma solução que, além de utilizar as estratégias disponibilizadas pelos autores no link do repositório citado anteriormente, incrementa algo a mais, de modo a melhorar a predição de similaridade entre diferentes músicas. Por exemplo, poderia-se utilizar a API do Spotify para pegar mais *features* relacionadas às músicas analisadas³, como por exemplo *loudness* e *danceability*, ou até mesmo uma API da Last.FM para pegar mais informações relevantes⁴, como por exemplo *tags* associada a uma certa música, de modo a conseguir computar com maior assertividade a similaridade entre diversos pares de músicas, além de possíveis alterações estruturais no algoritmo e novas abordagens de utilização e/ou exploração desses dados. Desse modo, esperamos poder alcançar um nível de predição de similaridade entre músicas melhor do que foi apresentado na solução escolhida pelo grupo anteriormente citado.

A base de dados não se encontra mais disponível em sua totalidade, pois o desafio já foi finalizado. Porém, uma boa parte desses dados se encontra disponível no repositório da solução⁵, podendo

ser utilizado livremente pelo projeto, desde que não haja interesse comercial nesses dados ou de aplicações geradas a partir deles. Esses dados contidos no projeto são arquivos de extensão JSON, contendo informações de *playlists* de usuários reais no sistema do Spotify (nome, músicas, autores, álbuns, visualizações, entre outras informações), totalizando, aproximadamente, 128 MB de dados, repartidos em 4 arquivos.

Para a métrica de avaliação será proposta uma medida relativa ao erro relativo da probabilidade da primeira música no ranking para ser recomendada e a música correta - entende-se, nesse caso, “correta” como a música que um usuário real inseriu em sua *playlist* -. Desse modo, conseguiremos quantificar a distância entre as predições e as respostas ideais, tendo uma ideia da acurácia do algoritmo proposto. Nesse caso, podemos definir o sucesso para a avaliação sendo uma procura para atingir a menor distância média entre todas as músicas e as predições de músicas recomendadas.

Portanto, ao final desse projeto de sistemas de recomendação, o grupo espera possuir um algoritmo que utilize de diferentes estratégias para recomendações e outras análises e abordagens desses dados. Isso se dará em função de que, esperançosamente, consigamos uma melhora gradativa das previsões de músicas a serem recomendadas para uma *playlist*.

¹<https://github.com/vaslnk/Spotify-Song-Recommendation-ML>

²<https://recsys-challenge.spotify.com>

³<https://developer.spotify.com/documentation/web-api/reference/tracks/get-several-audio-features>

⁴<https://www.last.fm/api/show/track.getInfo>

⁵<https://github.com/vaslnk/Spotify-Song-Recommendation-ML/tree/master/data>