

Recomendação de Músicas em Playlists no Spotify



Integrantes

Breno Martins
Marina Moreira
Ronald Pereira

1

Introdução



O Problema

- Recomendação de músicas
- Centenas de serviços de streaming
- Milhões de usuários
- Dezenas de milhões de músicas



O Problema

- Spotify RecSys Challenge 2018
- Participantes deveriam recomendar músicas para continuar uma playlist



Desafios

- Cold Start
- Popularity Bias
- Playlist generation/continuation



Desafios - Particularidades da música

- Duração
- Quantidade
- Consumo sequencial
- Comportamento do usuário
- Contexto



Estado da arte

- Combinação de filtragem colaborativa e extração de conteúdo
- Fatorização de matrizes
- Uso de técnicas de Machine Learning

2

Trabalhos Relacionados



Abordagens anteriores

- Filtragem colaborativa

Recomendação feita através de usuários ou itens similares

- Modelos baseados em contexto

Baseia-se em informações contextuais como comentários, reviews e tags



Abordagens anteriores

- Recomendações baseadas em metadados acústicos

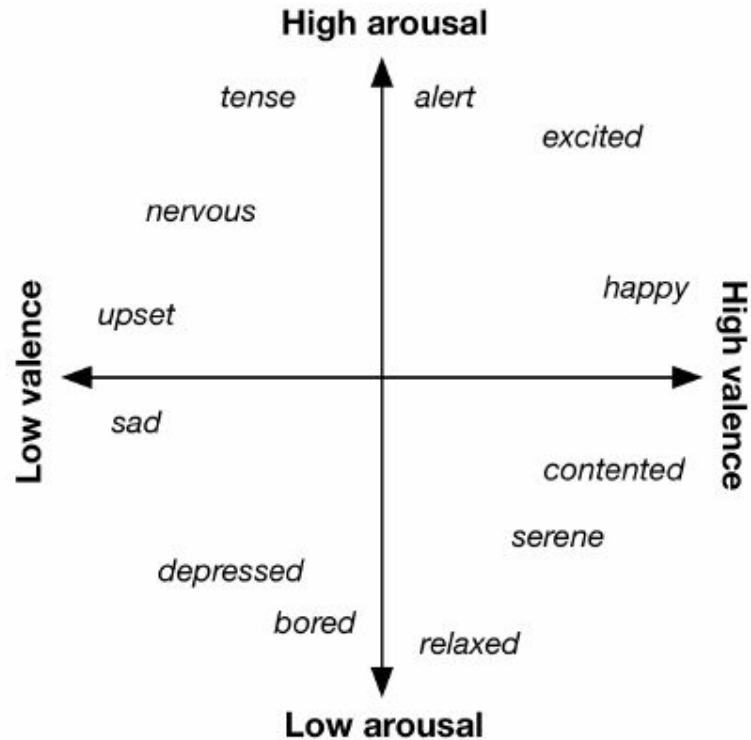
Utiliza propriedades acústicas através da análise de sinais de áudio

- Modelos baseados em emoção

Recomendação baseada no humor, onde o usuário registra suas emoções num espaço 2D de valence e arousal



Abordagens anteriores





Abordagens anteriores

- **Recomendações híbridas**

Combinação das abordagens citadas anteriormente



Relação com a solução proposta

- Será utilizada uma solução híbrida envolvendo filtragem colaborativa associada ao contexto e metadados acústicos
- Ao agregar diferentes *features* de uma música procura-se melhorar as predições

3

Solução Proposta e Descrição Experimental



Coleção de Treinamento/Teste

- Spotify Playlist Dataset
 - ▷ Conjunto de dados privado
 - ▷ 11.000 playlists
 - ▷ Colaborativas/Individuais
 - ▷ 731.360 músicas ao total
 - ▷ 180.409 músicas únicas
 - ▷ 37.476 artistas únicos



Coleção de Treinamento/Teste

	trackid	artist_name	track_name	pid	trackindex
0	spotify:track:0UaMYEvWZi0ZqiDOoHU3YI	Missy Elliott	Lose Control (feat. Ciara & Fat Man Scoop)	0	11647
1	spotify:track:6l9VzXrHxO9rA9A5euc8Ak	Britney Spears	Toxic	0	145808
2	spotify:track:0WqIKmW4BTj3eJFmnCKMv	Beyoncé	Crazy In Love	0	12511
3	spotify:track:1AWQoqb9bSvzTjaLralEkT	Justin Timberlake	Rock Your Body	0	27379
4	spotify:track:1lZr43nnXAijlGYnCT8M8H	Shaggy	It Wasn't Me	0	41255
5	spotify:track:0XUfyU2QviPAs6bxSpXYG4	Usher	Yeah!	0	12745
6	spotify:track:68vgtRHR7iZHpzGpon6Jlo	Usher	My Boo	0	142288
7	spotify:track:3BxWKCl06eQ5Od8TY2JBeA	The Pussycat Dolls	Buttons	0	73997
8	spotify:track:7H6ev70Weq6DdpZyyTmUXk	Destiny's Child	Say My Name	0	168512
9	spotify:track:2PpruBYCo4H7WOBj7Q2EwM	OutKast	Hey Ya! - Radio Mix / Club Mix	0	56109



Baselines



RecSys Challenge 2018

Final ranking for the main track

This leaderboard was computed on July 18, 2018 and shows the scores of final leaderboard for the **main** track. It takes into account the final submission scores evaluated against 100% of the challenge set, and rule compliance. Congratulations to the winners and to all that participated.

#	TEAM	RPREC	RPREC RANK	NDCG	NDCG RANK	CLICKS	CLICKS RANK	BORDA	CODE
1	vl6	0.22413802335	1	0.394622225334	1	1.7839	2	329	Code
2	hello world!	0.223353092072	2	0.393233606092	2	1.8952	6	323	Code
3	Avito	0.215316047292	6	0.384550795868	4	1.7818	1	322	Code



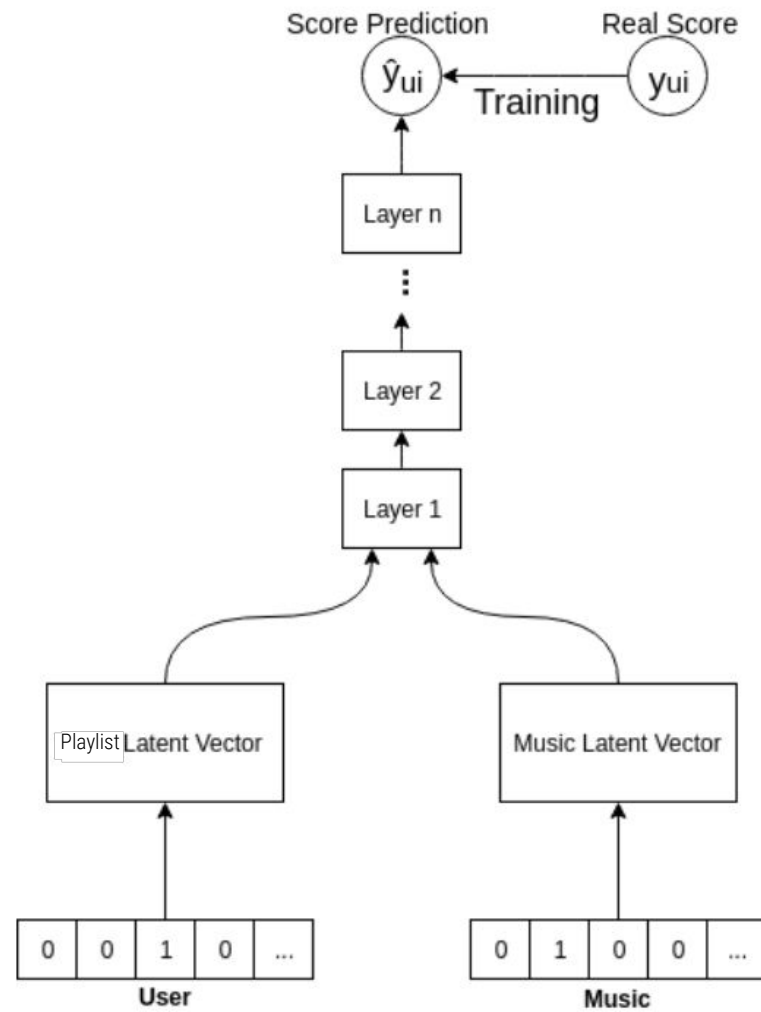
Baselines

- 1º lugar: V16
 - ▷ Feature Selection e XGBoost (XGBoost 4j)
 - ▷ Java
- 2º lugar: hello world!
 - ▷ Fatoração de Matriz e Predição de Ratings
 - ▷ Python 3 - Tensorflow para cálculos em GPU
- 3º lugar: Avito
 - ▷ Feature Selection e XGBoost
 - ▷ Python 3



Modelo Proposto

- Seleção de Features
- Fatorização de Matriz
 - ▷ Fatores Latentes de Playlists e Músicas
- Baseado em Conteúdo
 - ▷ API do Spotify
 - ▷ Loudness, Danceability, Energy, Tempo
- Rede Neural Densa
 - ▷ Keras + Tensorflow





Modelo Atual

Layer (type)	Output Shape	Param #	Connected to
playlist_input (InputLayer)	(None, 1)	0	
item_input (InputLayer)	(None, 1)	0	
playlist_embedding (Embedding)	(None, 1, 8)	88000	playlist_input[0][0]
item_embedding (Embedding)	(None, 1, 8)	1443272	item_input[0][0]
flatten_1 (Flatten)	(None, 8)	0	playlist_embedding[0][0]
flatten_2 (Flatten)	(None, 8)	0	item_embedding[0][0]
multiply_1 (Multiply)	(None, 8)	0	flatten_1[0][0] flatten_2[0][0]
prediction_layer (Dense)	(None, 1)	9	multiply_1[0][0]
Total params: 1,531,281			
Trainable params: 1,531,281			
Non-trainable params: 0			



Métricas de Desempenho

■ R-Precision

- ▷ $R_{prec} = |Músicas Relevantes Retornadas| / |Músicas Relevantes|$

■ Normalized Discounted Cumulative Gain (NDCG)

- ▷ Qualidade do ranking
 - ▷ Aumenta quando músicas relevantes estão no topo da lista

■ Recommended Songs Clicks

- ▷ Feature do Spotify
- ▷ Playlist → 10 músicas são recomendadas
 - ▷ Click → Outras 10 músicas são recomendadas
 - ▷ Mínimo de clicks para aparecer uma música relevante



Desempenho Atual

- Somente dados das playlists
 - ▷ Ainda não utiliza do conteúdo das músicas (API)
- Categoria Binária
 - ▷ Prediz se uma música tem que pertencer ou não à uma playlist
 - ▷ 80% de acurácia
 - ▷ $R_{prec} = 0.107374 = (0.8)^{10}$



OBRIGADO

Dúvidas?



github.com/ronaldpereira/spotify-playlist-song-recommendation/