

Introdução A Banco De Dados Recomendações de Filmes (Netflix)

Gabriel Souza e Silva
Gustavo Cordeiro Libel
Mateus Campaner Hercules

Universidade Tecnológica Federal do Paraná
Orientador: Prof. Dr. Luiz Celso Gomes Junior

28 de Novembro de 2017

Overview

- 1 Netflix Prize
- 2 Grafo de Correlações
 - Parâmetros Globais
 - Melhores Conexões
- 3 Recomendações
- 4 Validação
- 5 Continuação

Leaderboard



Leaderboard

10.05%

Display top leaders.

Rank	Team Name	Best Score	% Improvement	Last Submit Time
1	BellKor's Pragmatic Chaos	0.8558	10.05	2009-06-26 18:42:37
Grand Prize - RMSE \leq 0.8563				
2	PragmaticTheory	0.8582	9.80	2009-06-25 22:15:51
3	BellKor in BigChaos	0.8590	9.71	2009-05-13 08:14:09
4	Grand Prize Team	0.8593	9.68	2009-06-12 08:20:24
5	Dace	0.8604	9.56	2009-04-22 05:57:03
6	BigChaos	0.8613	9.47	2009-06-23 23:06:52

Leaderboard



Dados Netflix

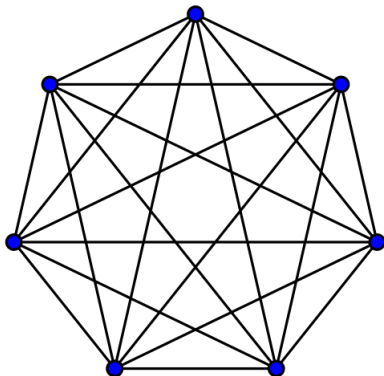
The screenshot shows a file explorer window with two tabs: 'movie_titles.txt' and 'mv_0000001.txt'. The 'movie_titles.txt' tab is active, displaying a list of movie titles and their corresponding user IDs. The 'mv_0000001.txt' tab is also visible, showing a list of user IDs and their corresponding movie ratings. A third window, 'NetFlix.csv', is open in the foreground, displaying a table with columns 'user', 'rate', and 'url'. The table contains data for various users and their ratings for different movies.

user	rate	url
0,3	0,062622	
0,3	1,1345836	
1,4	0,133093	
1,3	1,1345836	
1,3	0,062622	
2,5	0,457430	
2,2	1,1345836	
2,5	0,133093	

2.1GB → 3.9MB

≈ 250000 Avaliações

Grafo de Correlações



Vértice \rightarrow Filme

Aresta \rightarrow Peso & Confiabilidade

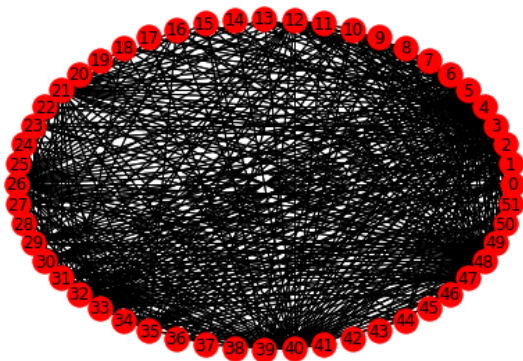
Parâmetros Globais

```
weight_by_rate=[[ 0, 0, 0, 0, 0],\
                 [ 0, 0, 0, 0, 0],\
                 [ 0, 0, 0, 1, 2],\
                 [ 0, -1, 1, 2, 3],\
                 [-1, 0, 2, 3, 4]]
min_reliability = 7
imdb_rate_relevance = 0.01 # 1%
```

A\B	1	2	3	4	5
1	0	0	0	0	0
2	0	0	0	0	0
3	0	0	0	1	2
4	0	-1	1	2	3
5	-1	0	2	3	4

Peso da aresta $A \rightarrow B$ em relação as notas dos filmes A e B dada por um mesmo usuário.

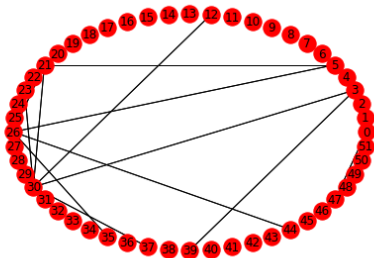
Grafo



Removidos:

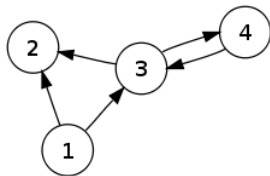
- Pesos = 0
- Confiabilidade < 7

Melhores Conexões



- Toy Story - Matrix
- Hotaru no haka - Monsters, Inc.
- The Matrix - A Beautiful Mind
- Toy Story - Monsters, Inc.
- Psycho - The Godfather
- The Matrix - WALL*E

Recomendações



Vértice \rightarrow Filme & Recomendação

Para cada Pessoa:

Para cada Avaliação:

Para cada Vizinho:

Atualiza Recomendação do Vértice

Ordena Vértices por Recomendação

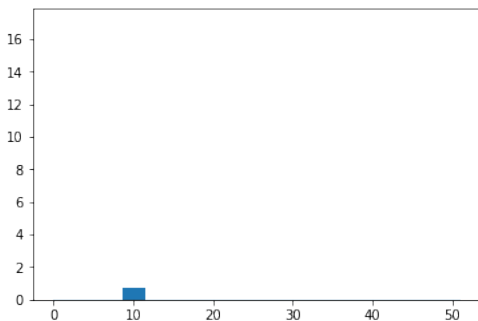
Recomendações

person	movie1	movie2	movie3	movie4	movie5
amandalima	The Matrix	Toy Story	Monsters, Inc.	Scarface	Il buono, il brutto, il cattivo
annabucharies	Monsters, Inc.	The Matrix	Il buono, il brutto, il cattivo	A Clockwork Orange	The Godfather
barbaragueros	The Matrix	The Dark Knight Rises	Dr. Strangelove or: How I Learned to Stop Worr...	El laberinto del fauno	Monsters, Inc.
fabiancajkowski	The Dark Knight Rises	2001: A Space Odyssey	El laberinto del fauno	Toy Story	Dr. Strangelove or: How I Learned to Stop Worr...
gabrielfernandes	Psycho	Kill Bill: Vol. 1	Toy Story	Monsters, Inc.	The Matrix
gabrielsilva	A Clockwork Orange	El laberinto del fauno	Toy Story	2001: A Space Odyssey	Memento
gustavolibel	The Matrix	Dr. Strangelove or: How I Learned to Stop Worr...	2001: A Space Odyssey	The Pianist	El laberinto del fauno
gustavopupo	Dr. Strangelove or: How I Learned to Stop Worr...	The Matrix	2001: A Space Odyssey	The Dark Knight Rises	El laberinto del fauno
igorlatini	The Matrix	A Clockwork Orange	El laberinto del fauno	2001: A Space Odyssey	The Pianist
katiensilva	2001: A Space Odyssey	Toy Story	The Dark Knight Rises	El laberinto del fauno	Dr. Strangelove or: How I Learned to Stop Worr...
lucasmatos	The Matrix	The Dark Knight Rises	2001: A Space Odyssey	Dr. Strangelove or: How I Learned to Stop Worr...	Monsters, Inc.
lucasseixas	Monsters, Inc.	Toy Story	The Matrix	Kill Bill: Vol. 1	The Dark Knight Rises

Recomendações

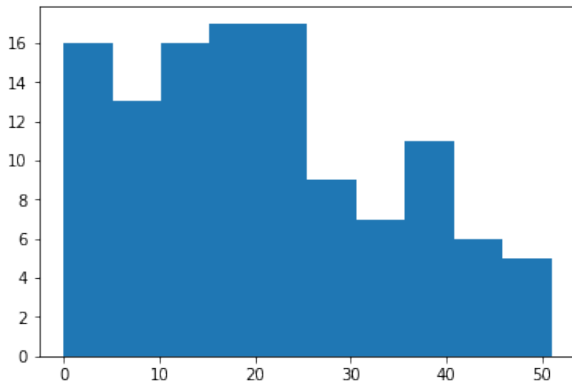
person	movie1	movie2	movie3	movie4	movie5
lucasseixas	Monsters, Inc.	Toy Story	The Matrix	Kill Bill: Vol. 1	The Dark Knight Rises
luispaiva	Il buono, il brutto, il cattivo	Toy Story	The Matrix	The Godfather	El laberinto del fauno
marcelomarciniak	The Matrix	The Dark Knight Rises	El laberinto del fauno	Toy Story	Kill Bill: Vol. 1
marcostholken	Toy Story	El laberinto del fauno	The Dark Knight Rises	2001: A Space Odyssey	Kill Bill: Vol. 1
mateushercules	The Dark Knight Rises	2001: A Space Odyssey	Toy Story	El laberinto del fauno	Monsters, Inc.
matheusgoncalves	Toy Story	El laberinto del fauno	The Dark Knight Rises	Kill Bill: Vol. 1	2001: A Space Odyssey
matheusgutierrez	The Dark Knight Rises	2001: A Space Odyssey	El laberinto del fauno	The Matrix	Dr. Strangelove or: How I Learned to Stop Worr...
natanreginaldo	El laberinto del fauno	2001: A Space Odyssey	The Dark Knight Rises	The Matrix	Dr. Strangelove or: How I Learned to Stop Worr...
otaviobastos	The Dark Knight Rises	2001: A Space Odyssey	El laberinto del fauno	Dr. Strangelove or: How I Learned to Stop Worr...	Kill Bill: Vol. 1
renatacarvalho	2001: A Space Odyssey	The Dark Knight Rises	El laberinto del fauno	Toy Story	A Clockwork Orange
rodrigossouza	Monsters, Inc.	A Clockwork Orange	The Matrix	Toy Story	The Dark Knight Rises
tomasbarreiros	The Matrix	Monsters, Inc.	2001: A Space Odyssey	The Dark Knight Rises	Dr. Strangelove or: How I Learned to Stop Worr...

Validação



$$\begin{array}{ccc}
 [4 & 23 & 44 & 16 & 42] \\
 [4 & 23 & 44 & 16] & [42] \rightarrow 10^\circ \\
 [4 & 23 & 44 & 42] & [16] \\
 \vdots & &
 \end{array}$$

Validação



media posições = 20.9

Continuação

```
weight_by_rate=[ [ 0, 0, 0, 0, 0], \
                  [ 0, 0, 0, 0, 0], \
                  [ 0, 0, 0, 1, 2], \
                  [ 0, -1, 1, 2, 3], \
                  [-1, 0, 2, 3, 4]]
min_reliability = 7
imdb_rate_relevance = 0.01 # 1%
```

$f(x)$ = media posições

$f : \mathbb{R}^n \rightarrow \mathbb{R}$

minimizar $f(x)$

Obrigado.