

# Sistemas de Recomendação

## Aplicando fatoração de matrizes em SRs

[R|adies]

10 de novembro de 2018



10 de novembro de 2018 1 / 51

(R|adies)

- 1 Introdução / Motivação
- 2 Conceito / Objetivo
- 3 Abordagens
- 4 SVD - Decomposição em Valores Singulares
- 5 Aplicando SVD em SRs
- 6 Amazon, Netflix, Spotify, Google, Youtube, etc



## Sumário

- 1 Introdução / Motivação
- 2 Conceito / Objetivo
- 3 Abordagens
- 4 SVD - Decomposição em Valores Singulares
- 5 Aplicando SVD em SRs
- 6 Amazon, Netflix, Spotify, Google, Youtube, etc

## Introdução

Para fazer uma escolha frente à tantas alternativas apresentadas, geralmente, confiamos nas opiniões passadas diretamente por outras pessoas, opiniões de revisores de filmes ou de livros, jornais impressos, mídia em geral, dentre outros.



Shardanand and Maes [1995]



(Radies)

10 de novembro de 2018 4 / 51

## Paradoxo da escolha

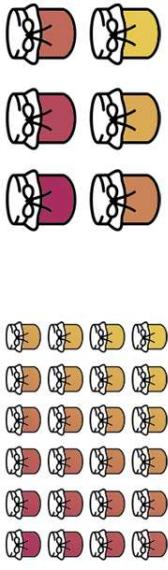
"À medida que aumenta o número de opções, o esforço exigido para tomar uma decisão acertada também aumenta; esse é um dos motivos pelos quais a escolha pode deixar de ser uma vantagem para se transformar em um ônus" (pag 68)



Schwartz [2004] - The Paradox of Choice: Why More is Less. Ecco, 1 edition.

## Paradoxo da escolha

### Too many choices?



### 24 choices of jam

attracted 60% of the shoppers  
3% of shoppers bought jam

### 6 choices of jam

attracted 40% of the shoppers  
**30%** of shoppers bought jam

Schwartz [2004] - The Paradox of Choice: Why More is Less. Ecco, 1 edition.



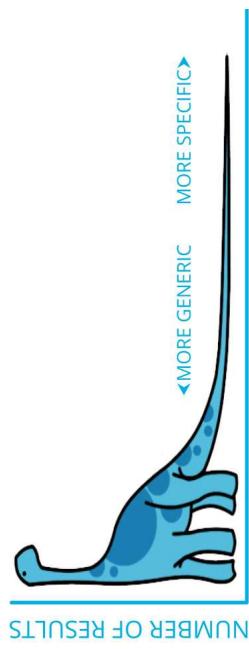
(Radies)

10 de novembro de 2018

6 / 51

## Cauda longa

O espantoso tamanho da Cauda Longa cria um mercado de produtos não-populares que rivaliza com o mercado dos populares.



Do mercado de massa para o mercado de nicho.

Anderson [2006] - The long tail. Random House Business.



## Cauda longa

- 25% das vendas da Amazon se situa fora dos 100 mil livros mais procurados.
- 21% das vendas do Netflix se concentram em filmes não populares.

Anderson [2006] - The long tail. Random House Business.

## Quem investe?

Os sistemas de recomendação são utilizados nas redes de comércio, como por exemplo, Amazon.com, eBay e Levis para estimular a compra de itens.

Schaffer et al [1999] - Recommender systems in e-commerce. In EC'99.  
Proceedings of the 1st ACM conference on Electronic commerce.



## Quem investe?

- Netflix Prize (busca de qualidade das recomendações)
- \$1million para quem melhorar em 10% o sistema de recomendação



(Radies)

10 de novembro de 2018 10 / 51

## Sumário

- 1 Introdução / Motivação
- 2 Conceito / Objetivo
- 3 Abordagens
- 4 SVD - Decomposição em Valores Singulares
- 5 Aplicando SVD em SRs
- 6 Amazon, Netflix, Spotify, Google, Youtube, etc

## Sistemas de recomendação

"São quaisquer sistemas que tenham por objetivo produzir como saída recomendações individualizadas, ou produzir o efeito de orientar o usuário de forma personalizada diante de uma grande variedade de opções."



Burke [2002] - Hybrid recommender systems: Survey and experiments. User modeling and User-Adapted Interaction.



## Objetivo

Auxiliar no processo natural e social de tomada de decisões, via automatização desse processo de indicação já há muito praticado nas relações humanas.



Ou seja, proporcionar ao usuário uma forma de identificação personalizada baseada nas suas preferências, aproximando-o do potencial conteúdo de interesse.

## Sumário

- 1 Introdução / Motivação
- 2 Conceito / Objetivo
- 3 Abordagens
- 4 SVD - Decomposição em Valores Singulares
- 5 Aplicando SVD em SRs
- 6 Amazon, Netflix, Spotify, Google, Youtube, etc

(Radies)

10 de novembro de 2018

14 / 51



## Sistemas de recomendação

Os sistemas de recomendação lidam com dois tipos de entidade: itens e usuários - e são geralmente classificados em três categorias:

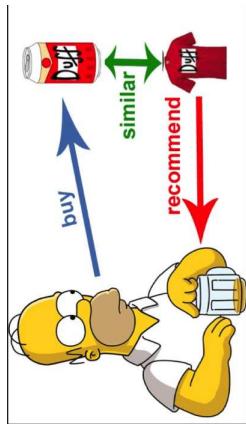
- Abordagem baseada em conteúdo,
- abordagem ou filtragem colaborativa e
- abordagem híbrida.

Member Adomavicius and Member-Tuzhilin [2005] - Toward the next generation of recommender systems: A survey of the state-of-the-art and possible extensions.



## Abordagem baseada em conteúdo

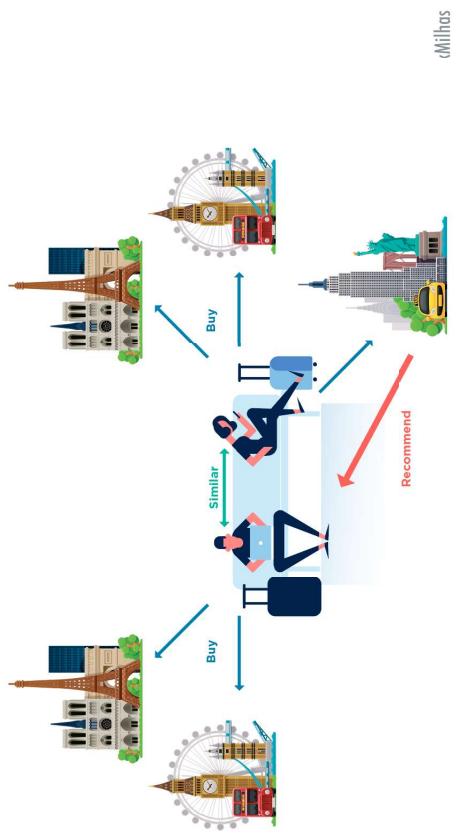
Abordagem onde são recomendados itens similares aqueles em que o usuário mostrou preferência no passado.



Abordagens

## Abordagem baseada em filtragem colaborativa

Nos sistemas colaborativos a essência está na troca de experiências entre as pessoas que possuem interesses comuns.

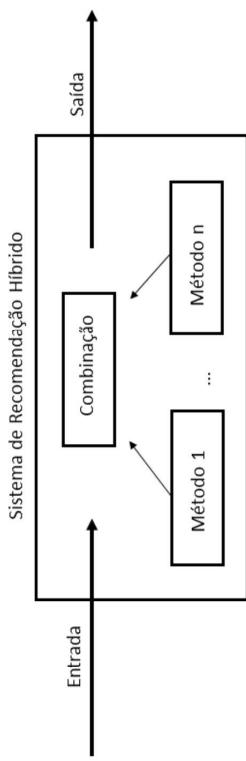


(Radians)

18 de novembro de 2018 17 / 51

1

## Abordagem híbrida



## Estratégia não personalizada - Top K

Geralmente, sugerem itens mais vendidos.

Problema: Cold Start = itens novos nunca foram avaliados por nenhum usuário ou usuários novos nunca avaliaram nenhum item.

Falta de passado!



## Sumário

- 1 Introdução / Motivação
- 2 Conceito / Objetivo
- 3 Abordagens
- 4 SVD - Decomposição em Valores Singulares
- 5 Aplicando SVD em SRs
- 6 Amazon, Netflix, Spotify, Google, Youtube, etc

(Radies)

10 de novembro de 2018

20 / 51



## Exemplo: Decomposição em Valores Singulares

$$\begin{matrix} n \\ m \end{matrix} \boxed{A} = \begin{matrix} n \\ m \end{matrix} \boxed{U} \begin{matrix} r \\ \Sigma \end{matrix} \boxed{V'}^n$$

Figura 3.2: SVD completa de uma matriz  $A_{n \times n}$  ( $m \geq n$ )

**Exemplo 3.1** [SVD completa.]

```
-> A = [1 1; 1 2; 1 3; 1 4];
-> [U,S,V] = svd(A)
V =
0.3220062 - 0.9467376
0.9467376 0.3220062
S =
5.7793788 0.
0. 0.7738091
0.
0.
0.
0.
U =
0.2195994 - 0.8073455 0.2236068 0.5
0.3833425 - 0.3912142 - 0.6708204 - 0.5
0.5471555 0.0249171 0.6708204 - 0.5
0.7109685 0.4410485 - 0.2236068 0.5
```



(Radies)

10 de novembro de 2018 21 / 51

## Exemplo: Decomposição em Valores Singulares

$$\begin{matrix} n \\ m \end{matrix} \boxed{A} = \begin{matrix} n \\ m \end{matrix} \boxed{U} \begin{matrix} r \\ \Sigma \end{matrix} \boxed{V^*} \begin{matrix} n \\ m \end{matrix}$$

Figura 3.2: SVD completa de uma matriz  $A_{m \times n}$  ( $m \geq n$ )

$$\begin{matrix} n \\ m \end{matrix} \boxed{A} = \begin{matrix} n \\ m \end{matrix} \boxed{U_1} \begin{matrix} r \\ \Delta \end{matrix} \boxed{V_1^*} \begin{matrix} n \\ r \end{matrix}$$

Figura 3.3: SVD reduzida de uma matriz  $A_{m \times n}$  ( $m \geq n$ ) (obs:  $\Delta = \Sigma$ )

## Exemplo: Decomposição em Valores Singulares

The screenshot shows the RStudio interface with two tabs open: 'exercicio.R' and 'exercicio.R'. The 'exercicio.R' tab contains the following R code:

```
1 require(Matrix)
2 
3 # 0, 3, 4, 0, 6
4 # 4, 1, 0, 0,
5 # 0, 0, 5, 0,
6 
7 data = read.csv(file="matrix.csv", header=FALSE, sep=",")
8 SA <- sparseMatrix(data$V1, data$V2, x=data$V3)
9 SVD <- svd(sa)
10 
11 SVD$u %*% diag(SVD$d)%*%t(SVD$v)
12 
13 (Top Level) ↴
```

The 'Console' tab shows the output of the code:

```
[1] -4.930630e-16 3.000000e+00 4.000000e+00 0 6.000000e+00
[2,] 4.000000e+00 1.000000e+00 9.860061e-16 0 -1.110223e-16
[3,] 7.771561e-16 1.110223e-15 5.000000e+00 0 1.000000e+00
```

A watermark 'MaxMilhas' is visible in the bottom right corner of the console area.

## Exemplo: Decomposição em Valores Singulares

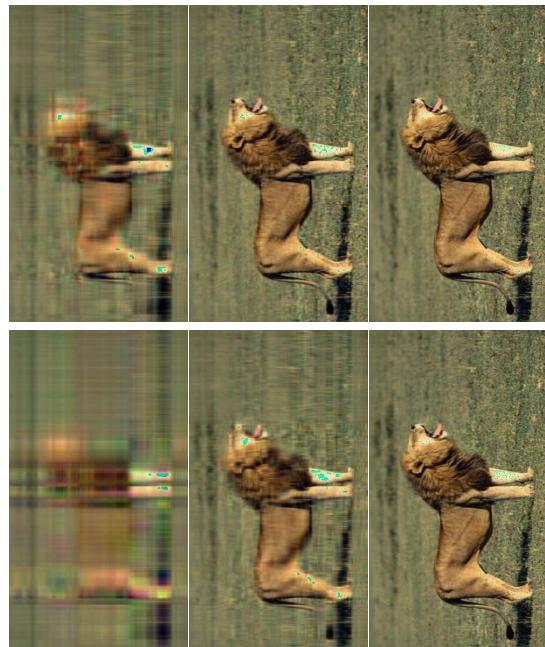
```

Console Terminal ✘
~/Documents/apresentacoes/Radies/ ↵
> SVD$u
[1] 0.4919325 [1] 0.112546837 [1] 0.08543528
[2] -0.4568336 [2] 0.283228299 [2] -0.28765317
[3] -0.4164226 [3] 0.385491771 [3] 0.24165159
[4] -0.3898678 [4] 0.085308989 [4] 0.2830478
[5] -0.2430259 [5] -0.125789082 [5] -0.48026513
[6] 0.2267956 [6] -0.245686583 [6] 0.7765925
[7] -0.3126194 [7] -0.825353294 [7] -0.14517913
[8] -0.1376715 [8] -0.633045048 [8] 0.05621762
> diag(SVD$u)
[1] 23.07781 [1] 23.07781 [1] 23.07781
[2] 0.00000 12.65567 0.0000000
[3] 0.00000 7.44694 0.0000000
> SVD$v
[1] -0.29121025 [1] -0.443406337 [1] 0.9190421
[2] -0.31299078 [2] -0.53562222 0.29108197
[3] -0.27257239 [3] 0.06962091 0.86614165
[4] -0.04488841 [4] -0.04236848 0.25044490
[5] 0.3135928 [5] 0.356322689 0.35504731
[6] -0.71289395 [6] 0.520653661 -0.13963230
[7] -0.19765419 [7] -0.13358990 -0.09961205
[8] -0.308822011 [8] 0.200653933 -0.09419710
4:15 (Top Level) :

```



## Compressão de Imagem: número de valores singulares



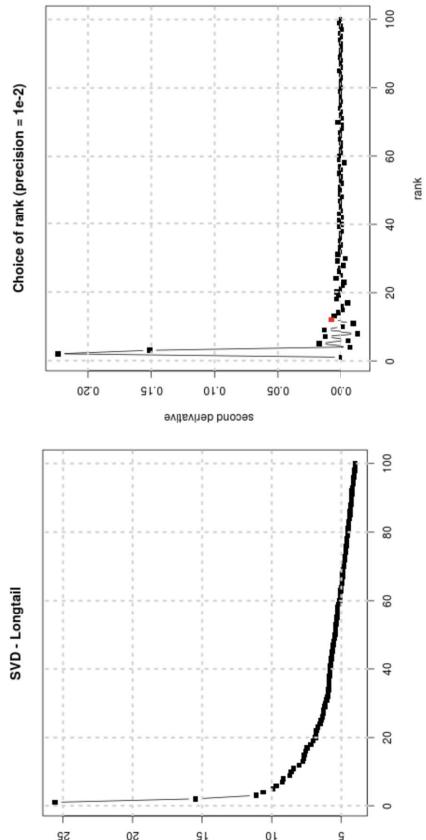
(Radies)

MaxMilhas

10 de novembro de 2018

25 / 51

## Escolha do número de valores singulares



## Sumário

- 1 Introdução / Motivação
- 2 Conceito / Objetivo
- 3 Abordagens
- 4 SVD - Decomposição em Valores Singulares
- 5 Aplicando SVD em SRs
- 6 Amazon, Netflix, Spotify, Google, Youtube, etc



## Modelagem = matriz de ratings

	Item 1	Item 2	Item 3	...	Item n
User 1	2	3	?	...	5
User 2	?	4	3	...	?
User 3	3	2	?	...	3
...	...	...	...	...	...
User m	1	?	5	...	4

## Aplicando SVD

### Example for SVD-based recommendation

• SVD:	$M_k = U_k \times \Sigma_k \times V_k^T$
$U_k$	Dim1 Dim2
Alice	0.47 -0.30
Bob	-0.44 0.23
Mary	0.70 -0.06
Sue	0.31 0.93
$V_k^T$	
Pretty Woman	0.38
Eat Prey Love	0.57
Twins	0.06
Die Hard	-0.57
Terminator	-0.44
$\Sigma_k$	Dim1 Dim2
Dim1	5.63 0
Dim2	0 3.23

• Prediction:  $\hat{r}_{ui} = \bar{r}_u + U_k(Alice) \times \Sigma_k \times V_k^T(EPL)$   
 $= 3 + 0.84 = 3.84$



## Aplicando SVD

```

exemplor.R x | Source on Save | Run | Source > | ↻
1 require(Matrix)
2 require(irriba)
3 raw.data <- read.table(file="data.st", header=FALSE, sep=',')
4 nusers = length(unique(raw.data$V1))
5 nitems = length(unique(raw.data$V2))
6 data <- sparseMatrix(i=raw.data$V1,
7 j=raw.data$V2,
8 x=raw.data$V3)
9 rsv = 5
10
11 rsv = 5
12
13 R <- irriba(A=data, nu=rsv, nv=rsv) # SVD procedure
14 d.adjust <- diag(sqrt(R$dd)) # square root of singular values (in a diagonal matrix)
15 U <- R$u %*% d.adjust # users features
16 V <- R$v %*% d.adjust # items features
17 recommendation = matrix(rnorm=nrow=rsv*nitems, ncol=3)
18
19 recommendation = matrix(rnorm=nrow=nusers*nitems, ncol=3)
20 k <- 1
21 for (User in 1:nusers){
22   REC <- U[User, ] %*% t(V)
23   for (Item in 1:nitems){
24     recommendation[k, 1] <- user
25     recommendation[k, 2] <- item
26     recommendation[k, 3] <- REC[item]
27     k <- k + 1
28   }
29 }
30

```

has

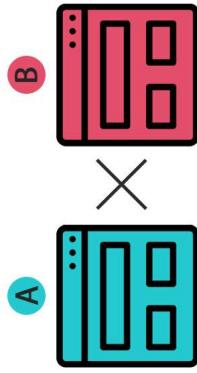
(Radies)

10 de novembro de 2018 30 / 51

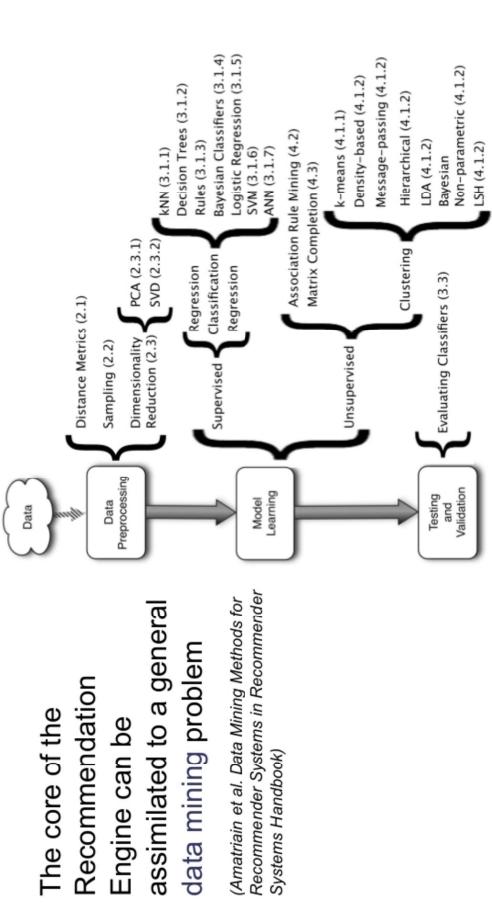
## Validação

$$\bullet \text{ offline} = \frac{\text{Exemplo: MAE e RMSE}}{\frac{\text{MAE} = \frac{1}{n} \sum_{j=1}^n |y_j - \hat{y}_j|}{\text{RMSE} = \sqrt{\frac{1}{n} \sum_{j=1}^n (y_j - \hat{y}_j)^2}}}$$

$$\bullet \text{ online} = \text{Teste A/B}$$



## Recomendação como Data Mining



## Sumário

- 1 Introdução / Motivação
- 2 Conceito / Objetivo
- 3 Abordagens
- 4 SVD - Decomposição em Valores Singulares
- 5 Aplicando SVD em SRs
- 6 Amazon, Netflix, Spotify, Google, Youtube, etc

Amazon, Netflix, Spotify, Google, YouTube, etc

*Amazon.com*

35% das vendas da Amazon vem de recomendações!!



(Radies)

10 de novembro de 2018

34 / 51

Amazon, Netflix, Spotify, Google, YouTube, etc

## Amazon.com

Gifts as unique as your mom Shop all gifts for Mother's Day

CARD

HEALTH & BEAUTY

ELECTRONICS

MOM-MADE FINDS

GIFTS UNDER \$30

PERSONALIZED GIFTS

Deals recommended for you See all deals

\$58.73 - \$67.87 Endt in 14:05:51

\$6.00 - \$799.99 Endt in 14:25:50

\$44.99 \$49.99 \$69.99 \$119.99

Endt in 14:25:50

MaxMilhas

10 de novembro de 2018 35 / 51

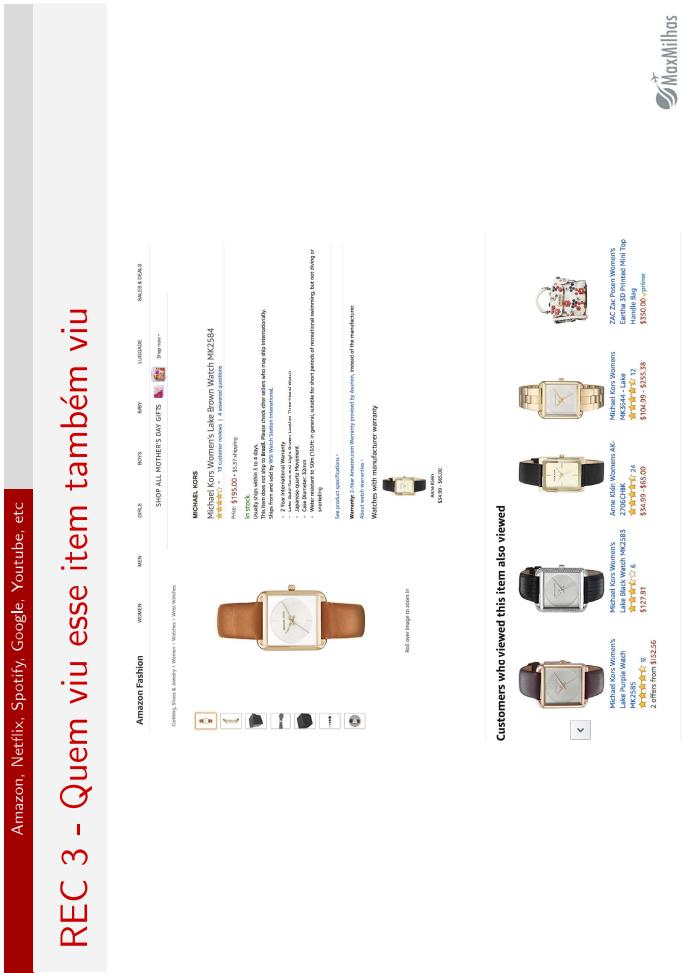
## REC 1 - Itens relacionados a esse item

**REC 2 - Quem comprou esse item também comprou**

The screenshot shows a product page for a Michael Kors Watch (MK2534) on Amazon Fashion. The watch has a gold-toned case and a white dial with black markers. It is paired with a brown leather strap. The page includes a 'Buy' button, a 'Buy Now' button, and a 'Add to Cart' button. A detailed product description highlights its features: it is a quartz movement, water-resistant up to 30 meters, and has a date window at the 3 o'clock position. The watch is described as a 'versatile timepiece' suitable for both casual and dressy occasions. Below the main product image, there is a 'Related items' section titled 'Customers who bought this item also bought'. This section displays five other watches, each with a small image, the brand name, model number, price, and a star rating. The watches include models from brands like Bulova, Akribos XXIV, and Anne Klein.

Amazon, Netflix, Spotify, Google, Youtube, etc

**REC 3 - Quem viu esse item também viu**



10 de novembro de 2018 38 / 51

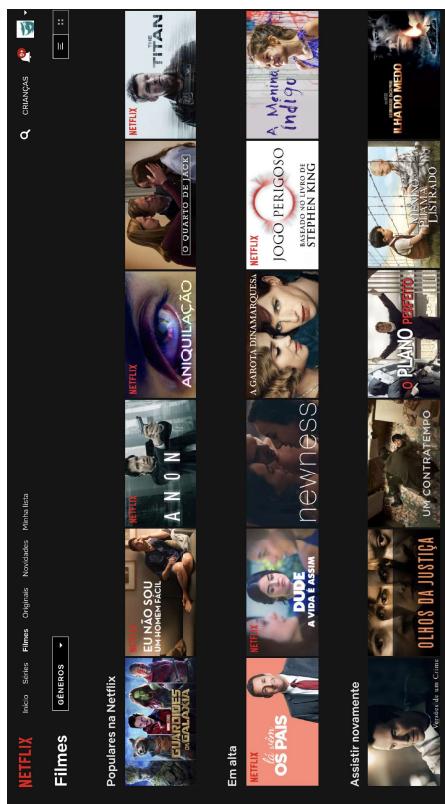
Radies

2/3 dos filmes vistos na Netflix vem de recomendações !!!



Amazon, Netflix, Spotify, Google, YouTube, etc

## Exemplo de recomendação no Netflix

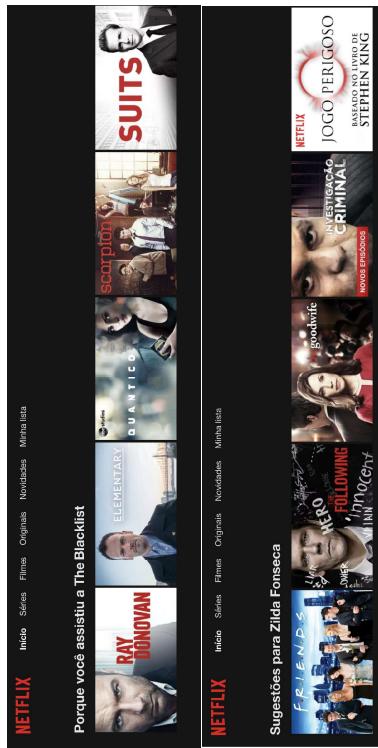


(Radies)

10 de novembro de 2018

40 / 51

## Exemplo de recomendação no Netflix



## Exemplo de recomendação no Spotify

**Navegar**

VÍDEO SESSIÃO PODCASTS PARADAS GÊNEROS E MOMENTOS LANÇAMENTOS DESCARAR SHOWS

**Playlists feitas só para você**

**Descobertas da Semana**  
Sua playlist semanal com músicas frequentadas, gêneros desenhados e playlists mais populares do seu Spotify.

**As melhores recomendações para você**

**Roda das Novidades**  
Não perca mais nenhuma lançamento! Faça o download das primeiras músicas desse setor novinho, mas já que é novo!

**PLAYLISTS | SÓ SPOTIFY**

**DEIXE-ME IR**  
Will Kulu & Praggo Waka

**ORIENTE**  
Yin-Yang

**SUNSHINE**  
Sunshine, Chico Franco & Zé Lelé

**A Vida É Chão**  
Dani Black

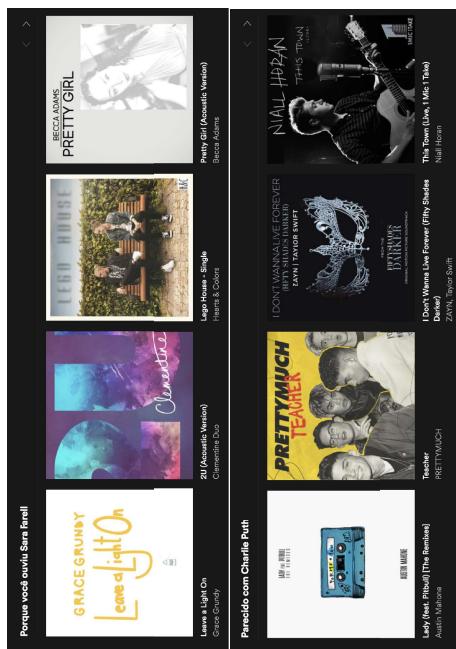
**MAX MILHAS**

10 de novembro de 2018 42 / 51

(Radies)

Amazon, Netflix, Spotify, Google, YouTube, etc

## Exemplo de recomendação no Spotify



10 de novembro de 2018

43 / 51

(Radies)

## Exemplo de recomendação na Google

38% das notícias clicadas no Google News vêm de recomendações!!



## Exemplo de recomendação na Google

The screenshot shows the Google News homepage with the following layout:

- Top Bar:** Includes a search bar with placeholder "Pesquisar", a language selector set to "U.S.", and a "MaxMilhas" logo.
- Left Sidebar:** Titled "SEÇÕES", it lists "Notícias principais" and various news categories: Mundo, Brasil, Negócios, Ciência/Tecnologia, Entretenimento, Esportes, and Saúde. Each category has a small icon and a link to its respective section.
- Main Content Area:** Titled "Notícias principais". It displays several news cards:
  - Morales envia investigação sobre Áécio Neves para Justiça de Minas**: Includes a thumbnail of a man, a summary, and links to "Galileo.com" and "Mais Informações Sobre".
  - Áécio Neves**: Includes a thumbnail of a man, a summary, and links to "Estado de Minas" and "RT.COM (licenciado de impresso)".
  - Sistema judicial Federal**: Includes a thumbnail of a man, a summary, and a "Ver cobertura completa" link.
  - Incerteza de Trump afeta a economia do rá**: Includes a thumbnail of a man, a summary, and a "EXAME.com" link.
  - Familias desabrigadas por desabamento de prédio em SP começam a sacar auxílio-moradia**: Includes a thumbnail of a group of people, a summary, and a "G1.GLOBO.COM" link.
  - Pré-candidatos Jair Bolsonaro e Ciro Gomes se reúnem para aliança**: Includes a thumbnail of two men, a summary, and links to "Datafolha" and "COBERTURA RELACIONADA".
  - 3 de maio de 2018 - Twitter Mais cidades**: Includes a thumbnail of a person, a summary, and a "Twitter" link.
- Bottom Bar:** A red bar containing the text "(Radies)".
- Page Footer:** A dark grey bar with the text "10 de novembro de 2018 45 / 51".

Amazon, Netflix, Spotify, Google, YouTube, etc

## Exemplo de recomendação no YouTube

The screenshot shows the YouTube mobile application interface. At the top, there's a banner for Claro 4.5G. Below it, the main content area displays several recommended video thumbnails. The first video in the list is 'LIBRADAIS TRACAS' by 'LIBRADAISS TRACAS'. Other recommended videos include 'Aventura do Rio - So Chær' by 'So Chær', 'MC MURKIN & MC LARANJA - Vai Vai' by 'MC MURKIN & MC LARANJA', 'Casa Bonita' by 'Casa Bonita', 'Nave e Oceano' by '2B Nave e Oceano', 'Fazendo 28' by 'Fazendo 28', 'Fazendo 29' by 'Fazendo 29', 'Fazendo 30' by 'Fazendo 30', 'Fazendo 31' by 'Fazendo 31', and 'Fazendo 32' by 'Fazendo 32'. The bottom of the screen features a navigation bar with icons for Home, Subscriptions, Library, History, and Account settings.

(Radies)

10 de novembro de 2018

46 / 51

## Resultado?

- 2/3 dos filmes vistos na Netflix vem de recomendações;
- 38% das notícias clicadas no Google News vem de recomendações;
- 35% das vendas da Amazon vem de recomendações.

# Matrix factorization

Amazon, Netflix, Spotify, Google, YouTube, etc

## Application of Dimensionality Reduction in Recommender System -- A Case Study

Badrul M. Sarwar, George Karypis, Joseph A. Konstan, John T. Riedl

GroupLens Research Group / Army HPC Research Center  
Department of Computer Science and Engineering  
University of Minnesota  
Minneapolis, MN 55455  
+1 612 625-4002

{sarwar, karypis, konstan, riedl}@cs.umn.edu

### Abstract

We investigate the use of dimensionality reduction to improve performance for a new class of data analysis software called "recommender systems". Recommender systems apply knowledge discovery techniques to the problem of making product recommendations during a live customer interaction. These systems are achieving widespread success in E-commerce nowadays, especially with the advent of the Internet. The tremendous growth of customers and products poses three key challenges for recommender systems in the E-commerce domain. These are producing high quality recommendations, performing many recommendations per second for millions of customers and products, and achieving high coverage in the face of data sparsity. One successful recommender system technology is *collaborative filtering*, which works by matching customer preferences to other customers in making recommendations. Collaborative filtering has been shown to produce high quality recommendations, but

experience suggests that SVT has the potential to meet many of the challenges of recommender systems under certain conditions.

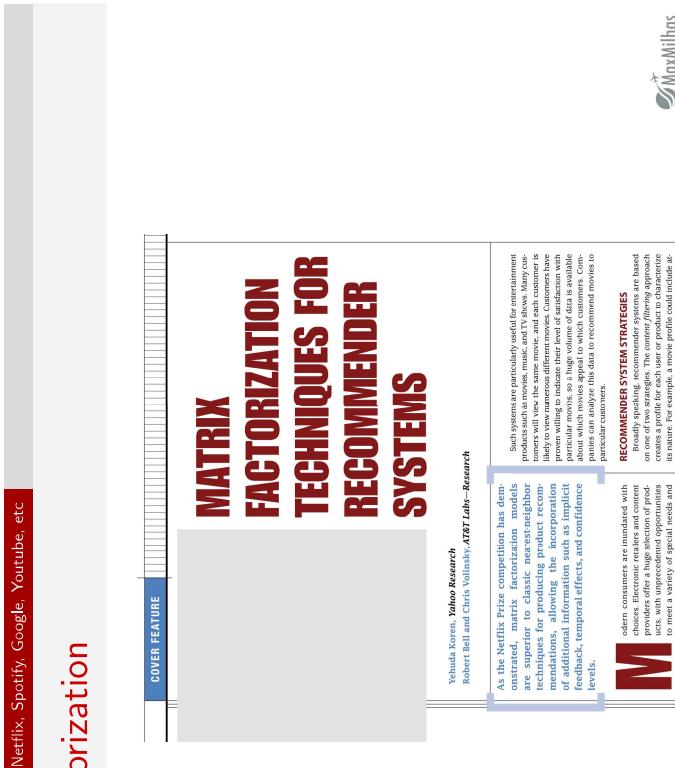
### 1 Introduction

Recommender systems have evolved in the extremely interactive environment of the Web. They apply data analysis techniques to the problem of helping customers find which products they would like to purchase at E-commerce sites. For instance, a recommender system on Amazon.com suggests books to customers based on other books the customers have told Amazon they like. Another recommender system on CDnow ([www.cdnow.com](http://www.cdnow.com)) helps customers choose CDs to purchase as gifts, based on other CDs the recipient has liked in the past. In a sense, recommender systems are an application of a particular type of Knowledge Discovery in Databases (KDD) ([Fayyad et al., 1996](#)) technique. KDD systems use many subtle data analysis techniques to achieve two unsolved goals. They are: *i*) to save



## Matrix factorization

Amazon, Netflix, Spotify, Google, YouTube, etc



(Radies)

10 de novembro de 2018

49 / 51

## Perguntas?