



Sistemas de Recomendação

Marlesson Santana



Sumário

1. Introdução
2. Métodos
3. Arquitetura e Frameworks
4. Hands On

Olá!



Sou o **Marlesson Santana**

- Pesquisador do **Deep Learning Brasil** - UFG
- Community Manager do **Data Hackers**
- CDO na **CQuantt**



CQuantt



Sou o **Arthur Lacerda**

- Graduado em **Ciência da Computação** - UFG
- Full-Stack e Cientista de Dados na **CQuantt**



bit.ly/2lIFyIn



bit.ly/2kyIAyJ

O que é um Sistema de Recomendação?

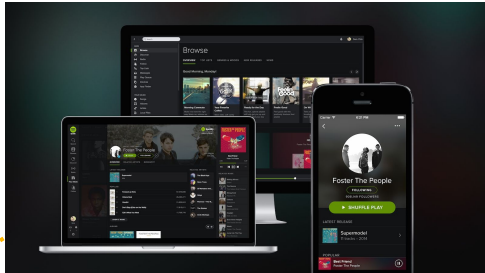
Introdução



*RecSys é o processo que usa as
preferências do usuário para
filtrar e sugerir conteúdo
personalizado.*

SPOTIFY

- 217 M de usuários
- 40 M de faixas
- 50 países
- 1.7 Bi



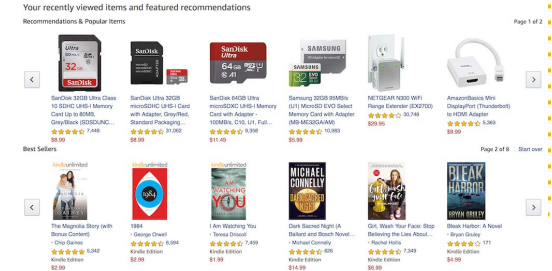
NETFLIX

- ~150 M de usuários
- ~10.000 títulos
- 190 Países
- **80% impactados com RecSys**



AMAZON

- 300 M de usuários
- 120 M de produtos
- 18 países
- **35% das compras vem de RecSys**





RecSys

- É aplicável em **diferentes domínios**
- Otimiza a **experiência do usuário** dentro da plataforma
- **Retém** o usuário dentro da plataforma
- **Revela padrões de consumo** que podem ser utilizados em outros setores

O **principal objetivo** de um RecSys é personalizar a experiência do usuário com foco em **aumentar o retorno financeiro**.

Famílias de Algoritmos

Métodos



Famílias de Algoritmos de RecSys



**Collaborative
Filtering**



**Content
Based**



Cold Start



O que são preferências do usuário?

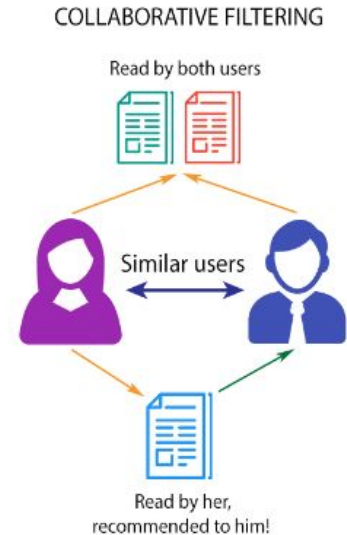
- Suas **interações** e/ou **notas** concedidas para determinado conteúdo
- Podem ser **explícitas** ou **implícitas**
- São **armazenadas** ao longo da utilização do sistema

Ex: curtidas, ratings, pause ou rewind, tempo de leitura, compartilhamentos.. etc






Collaborative Filtering

- Utiliza **preferências de outros usuários** para recomendar
- Necessita de muitos **dados históricos** de preferências
- Pode ser **computacionalmente custoso**





Collaborative Filtering

			
 Judite	10	?	7
 Moisés	8	9	6
 Rafael	1	3	10

Collaborative Filtering - Similaridade



similarity(👩, 👨) = similarity([10, ?, 7], [8, 9, 6])

- Euclidiana

$$\begin{aligned} d(\mathbf{p}, \mathbf{q}) &= d(\mathbf{q}, \mathbf{p}) = \sqrt{(q_1 - p_1)^2 + (q_2 - p_2)^2 + \dots + (q_n - p_n)^2} \\ &= \sqrt{\sum_{i=1}^n (q_i - p_i)^2}. \end{aligned}$$

- Cosseno

$$\text{similarity} = \cos(\theta) = \frac{\mathbf{A} \cdot \mathbf{B}}{\|\mathbf{A}\| \|\mathbf{B}\|} = \frac{\sum_{i=1}^n A_i B_i}{\sqrt{\sum_{i=1}^n A_i^2} \sqrt{\sum_{i=1}^n B_i^2}},$$

- Person

$$\begin{aligned} \text{Corr}(x, y) &= \frac{\sum_i (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum (x_i - \bar{x})^2} \sqrt{\sum (y_i - \bar{y})^2}} \\ &= \frac{\langle x - \bar{x}, y - \bar{y} \rangle}{\|x - \bar{x}\| \|y - \bar{y}\|} \\ &= \text{CosSim}(x - \bar{x}, y - \bar{y}) \end{aligned}$$

Collaborative Filtering - Similaridade



1



10

?

7

8

9

6

1

3

10

2

similarity(, )

3

1

0,9

0,6

1

0,6

1

4

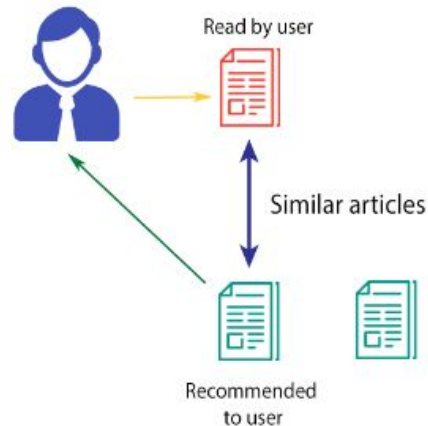
- Recomendar os **itens** que  consumiu e gostou, **mas**  ainda não viu.



Content Based




- Usa de **similaridade** entre conteúdos para recomendar
- Algoritmos de **rápida atualização**
- Necessidade de **modelagem de atributos** - o que é ser similar?
- Gera efeito de “**bolha de consumo**”

CONTENT-BASED FILTERING





Content Based

	Característica 1	Característica 2	Característica 3
	?	?	?
	?	?	?
	?	?	?

Content Based - Similaridade



1







2

similarity(, )

similarity(, )

3

			
	1	0,9	0,3

4

- Recomendar os **itens mais similares** que  **consumiu** e gostou



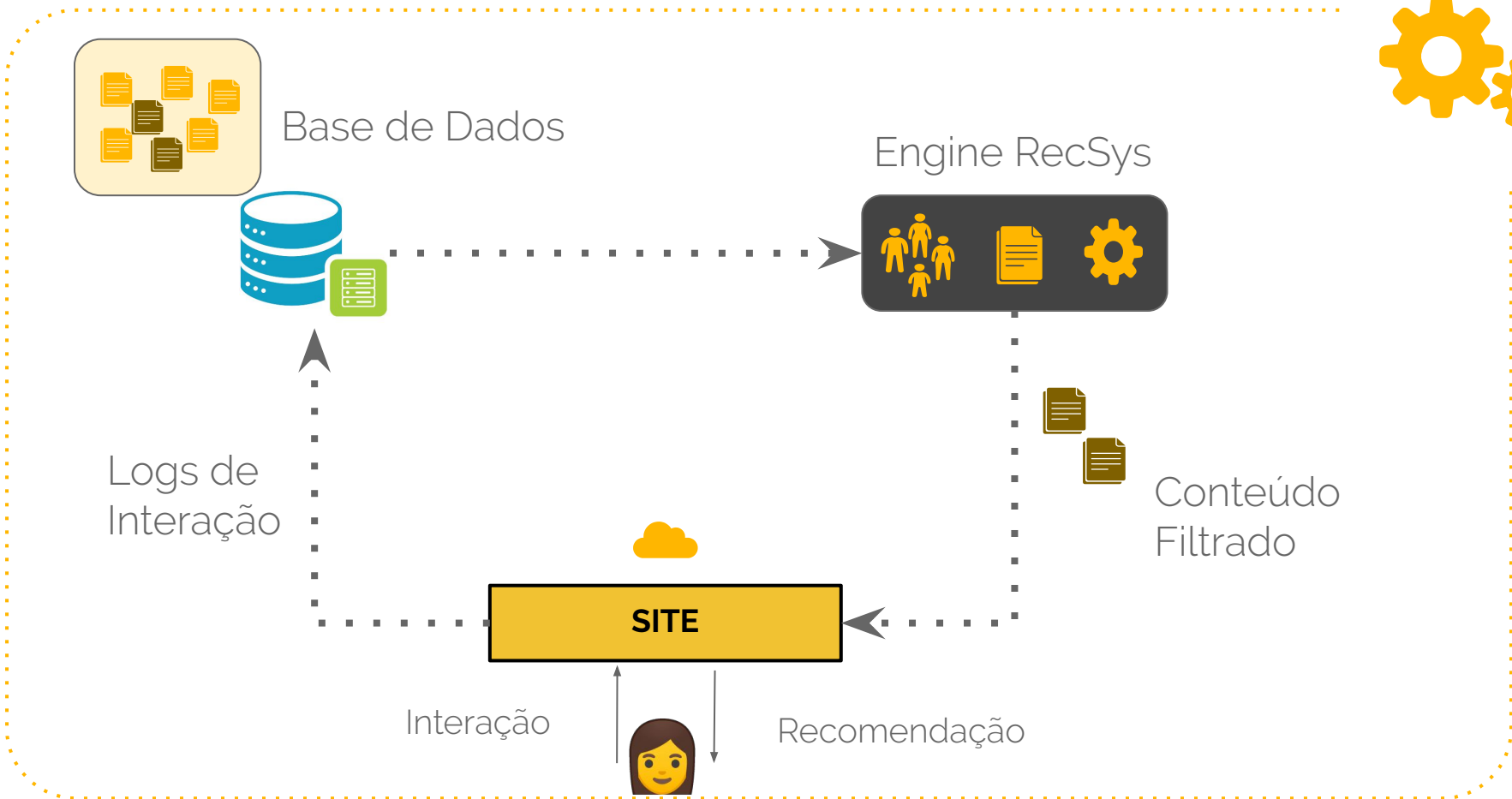
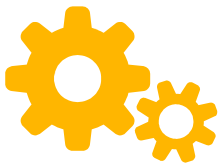
Cold Start

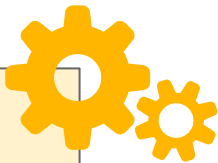
São listas de recomendação **sem ou com pouca personalização**.
Útil para novos usuários. Ex:

- Conteúdo "**mais acessados**"
- Conteúdo **geolocalizado**
- Listas personalizadas por **curadoria**

Arquitetura e Frameworks

Métodos





Coleta de Eventos



Armazenamento



Engine de RecSys



TensorFlow



Métricas de Avaliação

Métodos



Como avaliar a recomendação?

CTR

Taxa de Cliques. Se o que foi recomendado foi acessado.

Cobertura

Se todos os conteúdos da base estão sendo recomendados para alguém?

Diversidade

Se a recomendação está diversificada.

Engajamento

Se além de clicar o usuário engajou na recomendação (compartilhou, comentou.. etc).

Novidade

O quão nova é a recomendação.

Surpresa

Fator surpresa do que está sendo recomendado.

5

Hands on

Métodos

Obrigado!



Marlesson Santana

marlesson@cquantt.com

@marlesson



CQuantt