

Aprendizado de Máquina

Aula 9: Raciocínio Baseado em Casos e Sistemas de Recomendação (parte 1)

André C. P. L. F de Carvalho
ICMC/USP

andre@icmc.usp.br



Tópicos do módulo

- Introdução
- Raciocínio baseado em casos
- Sistemas de recomendação
- Principais abordagens
- Filtragem por conteúdo
- Filtragem colaborativa
- Métodos baseados em modelos
- Métodos baseados em memória

Tópicos do módulo

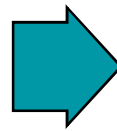
- Introdução
- Raciocínio baseado em casos
- Sistemas de recomendação
- Principais abordagens
- Filtragem por conteúdo
- Filtragem colaborativa
- Métodos baseados em modelos
- Métodos baseados em memória

Sistemas baseados em regras

- Muito usados no passado
 - Problema para construí-los:
 - Dificuldade de especialistas em traduzir experiência em regras



EXPERIÊNCIA



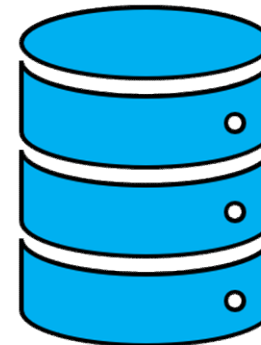
*If ...
Then ...
Else...*

REGRAS

Raciocínio baseado em casos



EXPERIÊNCIA



**BASE DE
EXPERIÊNCIAS**

Mas não uma BD!

Raciocínio baseado em casos

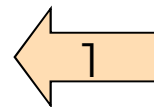
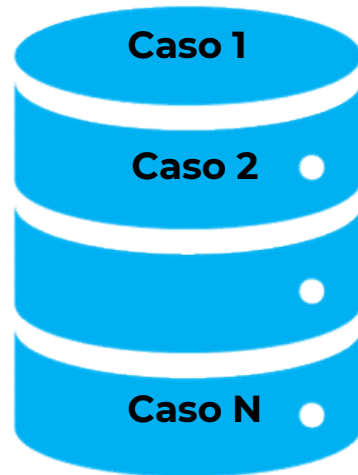
- Case-based reasoning (CBR)
- Mais que um sistema de gerenciamento de banco de dados
 - Sistemas de CBR procuram casos anteriores semelhantes ao problema (caso) atual
 - Buscam por eles em uma base de casos indexada
 - Indexação e representação de casos facilitam
 - Recuperação de casos relevantes e
 - Comparação desses casos com o problema atual
 - Permitem ainda a adaptação de casos recuperados

O que é um caso?

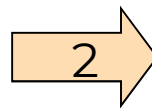
- Situação problemática, estrutura de dados com duas partes
 - Uma parte problema
 - Descrição de características (situações) de problemas específicos
 - Usada para identificar um caso
 - Indexação e recuperação
 - Uma parte solução
 - Explica como o caso foi resolvido anteriormente
 - De forma bem (ou mal) sucedida
 - Pode ser adaptada quando o caso é recuperado

Como funciona um sistema de CBR?

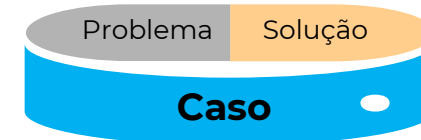
- Resolve novos problemas adaptando soluções usadas para casos anteriores que são parecidos



Descrição de um novo problema

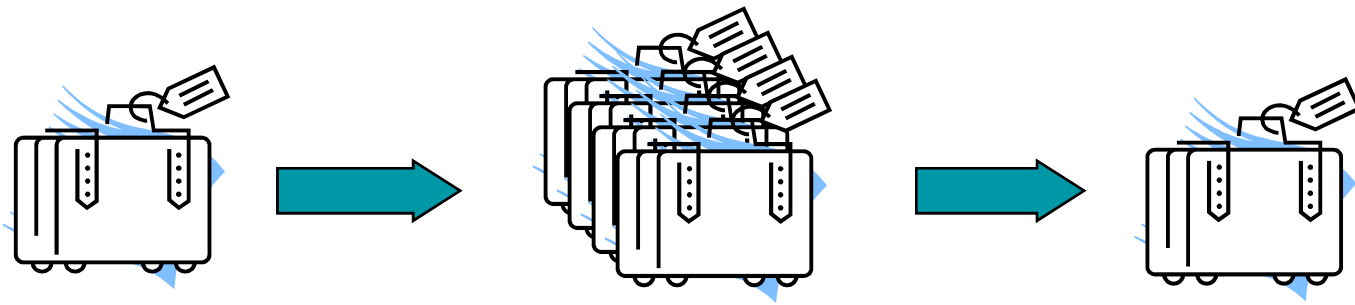


Retorna uma possível solução



Passos de um sistema de CBR

- Apresentar descrição do problema a ser resolvido
- Recuperar um ou mais casos com problema similar da base de casos
- Adaptar solução do(s) caso(s) recuperado(s)

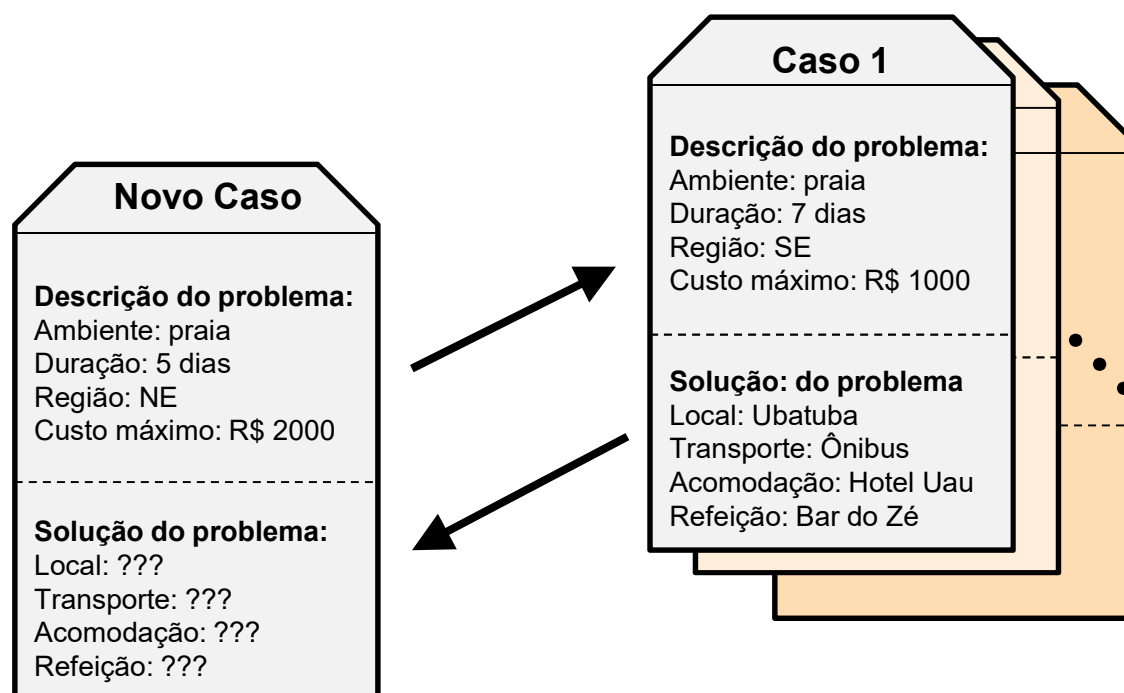


Problema: Que pacote de viagem comprar?

Casos (pacotes) com descrição de problema semelhante recuperados

Adaptação de solução de caso selecionado dentre os recuperados

Sistemas de CBR

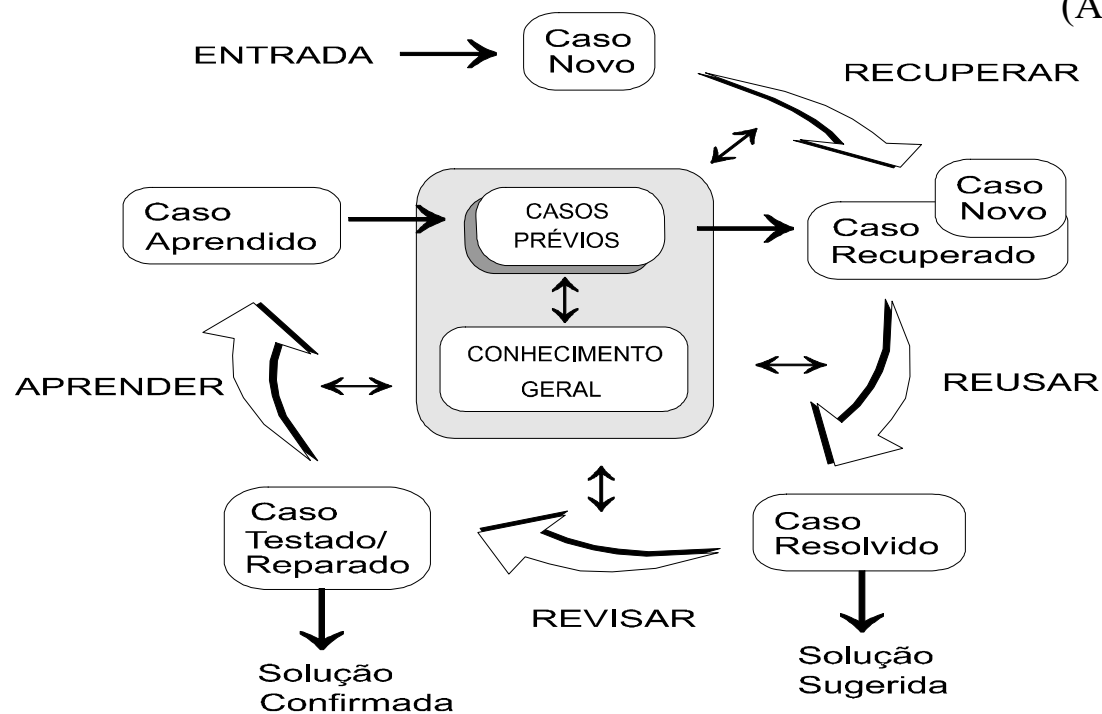


Sistemas de CBR

- Em geral, um sistema de CBR possui cinco processos:
 - Representação de casos
 - Indexação de casos
 - Armazenamento e recuperação de casos
 - Adaptação de casos
 - Avaliação e aprendizado de casos

Ciclo de um sistema de CBR

(Adaptado de Aamodt, 1993)

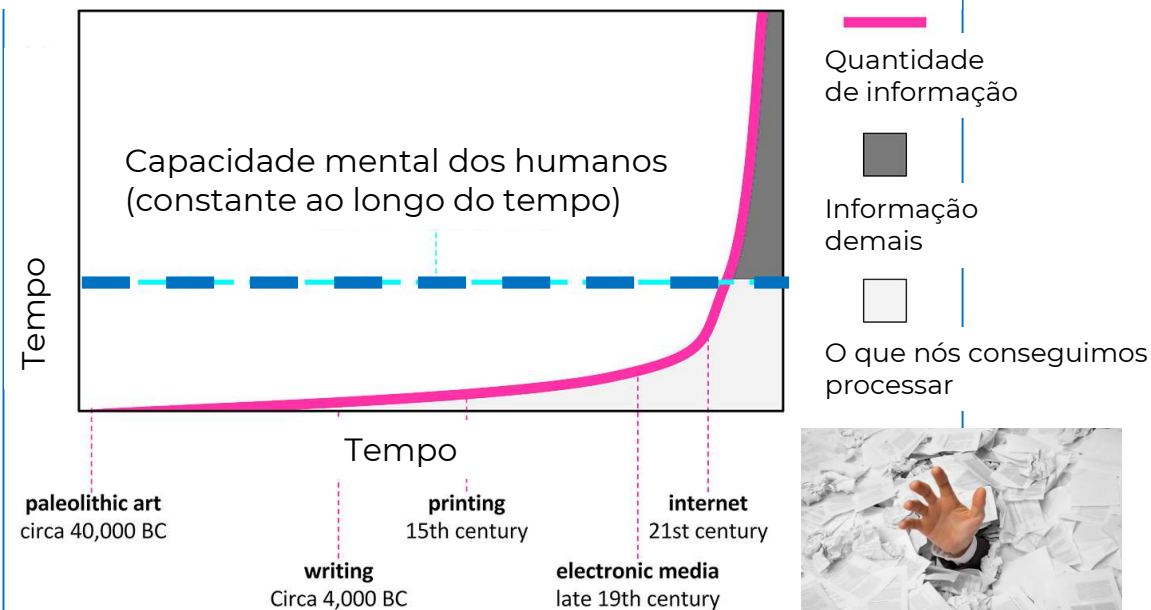


Sistemas de CBR

- Recuperação de casos previamente armazenados usa medida de similaridade
 - Importante passo do ciclo RBC
- Similaridade é geralmente medida por abordagens simples
 - Pode usar algoritmos de aprendizado de máquina
 - Ex.: algoritmo KNN
- Uma boa recuperação depende do uso de uma função de similaridade adequada
 - Frequentemente desconhecida

Sobrecarga de informação

Mais informação do que podemos processar



<https://pagefair.com/blog/2016/premium-publishers-trust-and-adblocking/ryans-loose-theory-of-tmi/>

Sobrecarga de informação

- Torna impossível para uma pessoa processar toda a informação disponível
 - Quanto mais opções nós temos, mais difícil fazer boas escolhas
 - Além de levar mais tempo e depender de conhecimento nos temas presentes da informação disponível
 - Necessidade de alternativas automatizadas para filtrar a informação disponível
 - Mantendo apenas o que for relevante, em quantidade gerenciável
 - Sistemas de recomendação

Netflix

- Fundada em 1997
 - Net (internet) Flix (Flicks, sinônimo de filmes)
 - Início: alugava DVDs de filmes, 1 ano depois enviava DVDs pelo correio
 - Vendia todos os filmes e shows de TV já feitos
- 2007
 - Netflix inicia mudança do seu modelo de negócios



Prêmio Netflix

- Outubro de 2006: lançou uma competição com o desafio:
 - Tornar seu algoritmo de recomendação 10% mais preciso
 - Reduzir $RMSE=0,9525$ para $RMSE<0,857$
 - Oferece US\$ 1.000.000,00 ao primeiro time que conseguir
 - Disponibilizou conjunto de dados com avaliações de 17.770 filmes por 480.189 clientes
 - Vinha tentando por anos melhorar o sistema com recrutamento de cientistas renomados

Prêmio Netflix

- 4º Colocado: Time **Simon Funk**
 - Submetida em julho de 2009
 - Adaptou sua abordagem para decomposição em valores singulares incremental
 - Não aumentou precisão em mais de 10%
- Vencedor: Time **BellKor Pragmatic Chaos**
 - Fatoração de matrizes
 - Time **The Ensemble** superou o resultado, mas 20 minutos depois
 - Prêmio foi doado para uma instituição de caridade

It is our great honor to announce the \$1M Grand Prize winner of the Netflix Prize contest as team [BellKor's Pragmatic Chaos](#) for their verified submission on July 26, 2009 at 18:18:28 UTC, achieving the winning RMSE of 0.8567 on the test subset. This represents a 10.06% improvement over Cinematch's score on the test subset at the start of the contest. We congratulate the team of Bob Bell, Martin Chabbert, Michael Jahrer, Yehuda Koren, Martin Piotte, Andreas TÅlscher and Chris Volinsky for their superb work advancing and integrating many significant techniques to achieve this result.

Sistemas de recomendação

- Recommender systems (RSs)
- Permitem lidar com sobrecarga de informação
 - Quando existe um grande número de opções para que um usuário ativo escolha uma delas
 - Usuário ativo: usuário que usa o RS para receber uma recomendação
 - Ajudando as pessoas a fazerem escolhas melhores
- Assim como algoritmos de aprendizado de máquina (AM), RSs utilizam álgebra linear
 - Manipulação de matrizes
 - Matrizes de avaliações (rating matrices)

Matriz de avaliações

- Rating Matrix (RM)
- Matriz Usuário-Item
 - U (usuários)
 - I (itens)
 - R_{ij} : avaliação (rating) do usuário i para o item j

		Itens (I)			
Usuários (U)	1	R_{11}	R_{12}	...	R_{1m}
	2	R_{21}			
			
			
	n	R_{n1}			
		1	2	...	m

Matriz de avaliações

- Banco de dados com avaliações do usuário ativo e de outros usuários

	Item1	Item2	Item3	Item4	Item5
Eu	5	3	4	?	7
Usuário1	3	1	2	3	3
Usuário2	4	3	4	3	5
Usuário3	3	3	1	5	4
Usuário4	1	5	5	2	1

Predição e recomendação

- Recomendação pode se considerada uma generalização da predição

- Conjunto de dados para tarefas preditivas

- Tabela atributo-valor

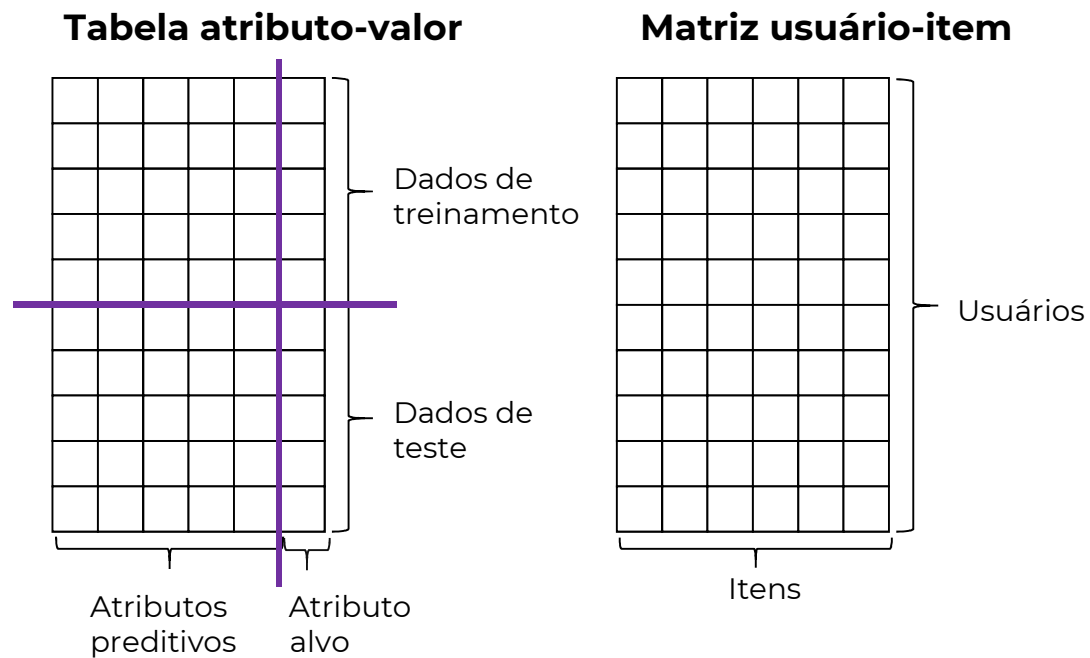
- Possui atributos preditivos e um atributo alvo
- Prediz valor para atributo alvo

- Conjunto de dados para um sistema de recomendação

- Matriz de avaliações

- Não existe distinção entre atributos preditivos e alvo
 - Qualquer atributo pode ser preditivo ou alvo
 - Pode prever valores para qualquer coluna da matriz de avaliações
- Prediz valor (avaliação) para item alvo

Predição e recomendação



Predição e recomendação

- Em geral, algoritmos de AM são aplicados a tabelas atributo-valor completas
 - Conjuntos de dados estruturados com todos os valores preenchidos
 - Caso existam valores ausentes
 - Como alguns algoritmos de AM não funcionam com valores ausentes, eles devem ser estimados antes
- Matrizes de avaliações são geralmente incompletas
 - Grande quantidade de valores ausentes

Matriz de avaliações

- Rating Matrix (RM)
- Matriz Usuário-Item
 - U (usuários)
 - I (itens)
 - R_{ij} : avaliação do usuário i para o item j
 - Ausência de feedback: ?

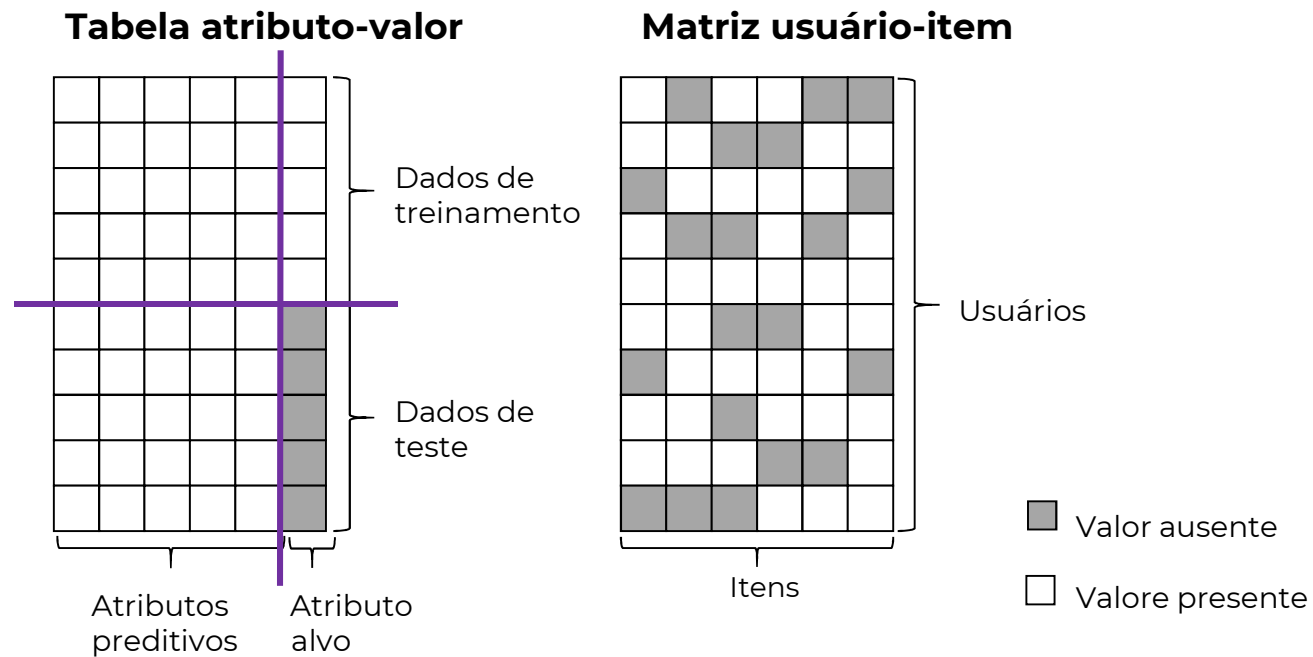
		Itens (I)			
Usuários (U)	1	R_{11}	R_{12}		?
	2	?	R_{22}		R_{2m}
	:	...		:	
	:				
	n	?	?		R_{nm}
		1	2	...	m

Matriz de avaliações

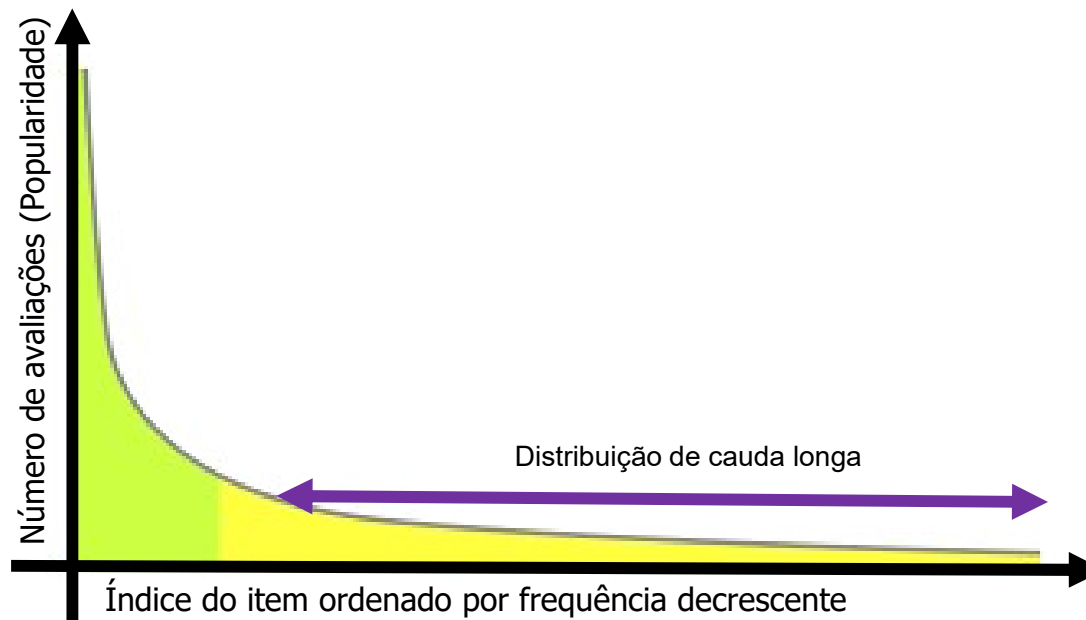
- Banco de dados com avaliações do usuário ativo e de outros usuários

	Item1	Item2	Item3	Item4	Item5
Eu	5	3	4	?	7
Usuário1	3	?	2	3	3
Usuário2	?	?	4	3	?
Usuário3	3	3	?	5	4
Usuário4	1	?	?	?	1

Predição e recomendação



Frequência de avaliações



Sistemas de recomendação

- Exemplos de produtos recomendados por RSs

Empresa que usa sistema de recomendação	Produto recomendado
Amazon	Vários tipos
Facebook	Amigos, notícias e propagandas
Google	Páginas
Google news	Notícias
Jester	Piadas
Netflix	Filmes e séries
Spotify	Músicas
Trip advisor	Pacotes de viagens

Contínua na
próxima aula