



COPPE
UFRJ

Instituto Alberto Luiz Coimbra de
Pós-Graduação e Pesquisa de Engenharia

***Uma Introdução à Filtragem
Colaborativa para Sistemas de
Recomendação***

Geraldo Zimbrão
Coppe/IM-DCC/UFRJ

Motivação

- Perspectiva do Usuário
 - Muitos produtos, livros, filmes, músicas etc.
 - Me ajude a escolher ... por favor!
 - Perspectiva do gerente
 - *"Se eu tenho 3 milhões de **clientes** na web, devo ter 3 milhões de lojas na web".*
- Jeff Bezos, CEO da Amazon.com*

**oferta
do dia**

ofertas exclusivas pra você :)
corre que é **só hoje!**

estoque limitado

13 h 31 m 40 s

3% de desconto



ar condicionado split hi wall electrolux
ecoturbo 9000 btus frio r410

★★★★★ (122)

R\$ 1.094,16

em 1x no cartão Americanas ou R\$ 1.128,00
em 12x de R\$ 94,00 sem juros

Receba R\$ 21,89 (2% de volta)

50% de cashback



prancha ga.ma eleganza plus bivolt
preta

★★★★★ (16)

R\$ 51,99

5x de R\$ 10,39 sem juros com Ame

Receba R\$ 26,00 (50% de volta)

frete grátis



xiaomi airdots fones de ouvido
esportivos bluetooth com handsfree

~~R\$ 206,10~~

R\$ 189,26

2x de R\$ 94,63 sem juros

Receba R\$ 3,79 (2% de volta)

internacional

retire hoje na loja mais próxima!



liquidificador mondial premium I-1000R
3l 12 velocidades vermelho - 1000w

★★★★★ (29)

R\$ 119,90

5x de R\$ 23,98 sem juros

Amazon.com.br

Amazon.com.br eBooks Kindle: A guerra do lobo - Crônicas Saxônicas - vol. 11, Bernard Cornwell - Google Chrome

Amazon.com.br eBooks Kindle: A guerra do lobo - Crônicas Saxônicas - vol. 11, Bernard Cornwell - Google Chrome

Olá, Geraldo
Contas e Listas Pedidos Seu Prime Carrinho

Enviar para Gracilene Orlando 32837

Comprar novamente Geraldo, sua Amazon.com.br Ofertas do Dia Venda na Amazon Ajuda Computadores Casa Livros Eletrônicos

Loja Kindle Promoções Kindle Kindle Unlimited Prime Reading Comprar Kindle Acessórios App Kindle eBooks Mais Vendidos Novidades Volta às aulas Dispositivos e conteúdo Ajuda

Volta aos resultados

A guerra do lobo - Crônicas Saxônicas - vol. 11 eBook Kindle
por Bernard Cornwell (Autor)

★★★★★ 64 avaliações de clientes

> Ver todos os 2 formatos e edições

Kindle R\$ 22,39	Capa Comum R\$31,99 prime
Leia com nossos apps gratuitos	1 Usado(s) a partir de R\$39,90 15 Novo(s) a partir de R\$31,99

Na luta pelo poder, dois inimigos mortais vão se encontrar frente a frente no aguardado novo volume de *Crônicas Saxônicas*.

Lorde Uhtred de Bebbanburg sabe que a paz ainda é um sonho distante. Embora tenha vencido a batalha pelo seu lar ancestral, rebeliões ameaçam a Mércia e os noruegueses espreitam em busca de oportunidades de invadir as terras

< Leia mais

Número de páginas: 378 páginas
Idioma: Português

Configuração de fonte: Habilitado
Page Flip: Habilitado

eBooks em oferta na Loja Kindle
eBooks promocionados por 24h: [inscreva-se](#) e não perca nenhuma oferta.

Preço digital sugerido: ~~R\$ 37,90~~
Preço Kindle: **R\$ 22,39**
Economize **R\$ 15,51 (41%)**

(Crédito) Visa ****1114 [Alterar](#)

[Comprar agora com 1-clique](#)

Enviar para:
Kindle Simple

Envie uma amostra grátis

Enviar para:
Kindle Simple

[Adicionar à Lista](#)

[Insira um cupom de desconto](#)

[Compartilhar](#) [Enviar](#) [Facebook](#) [Twitter](#) [Pinterest](#) [Incorporar](#)

Clientes que visualizaram este item também visualizaram

Página 1 de 8



Amazon.com

Your Amazon.com - Google Chrome

Your Amazon.com

https://www.amazon.com/gp/yourstore/home?ref=nav_cs_ys

Gmail Google Agenda Importantes JABOT VDO COPPETEC HELP postgre... Google Google Maps MQL5 Site Other bookmarks

Amazon Try Prime

All

EN Hello, Geraldo Account & Lists Orders Try Prime Cart

Deliver to Gracilene Orlando 32837

Whole Foods Browsing History Geraldo's Amazon.com Today's Deals Buy Again Gift Cards Help Registry Sell

Your Amazon.com Your Browsing History Recommended For You Improve Your Recommendations Your Profile Learn More

Recommended for you, Geraldo



Buy it again
10 ITEMS



New Releases
12 ITEMS



Health Care Products
21 ITEMS



Vitamins & Dietary Supplements
84 ITEMS



Screen Protector
QHOHQ



Memory Stick PRO Duo
compatible
micro SD
Adaptor

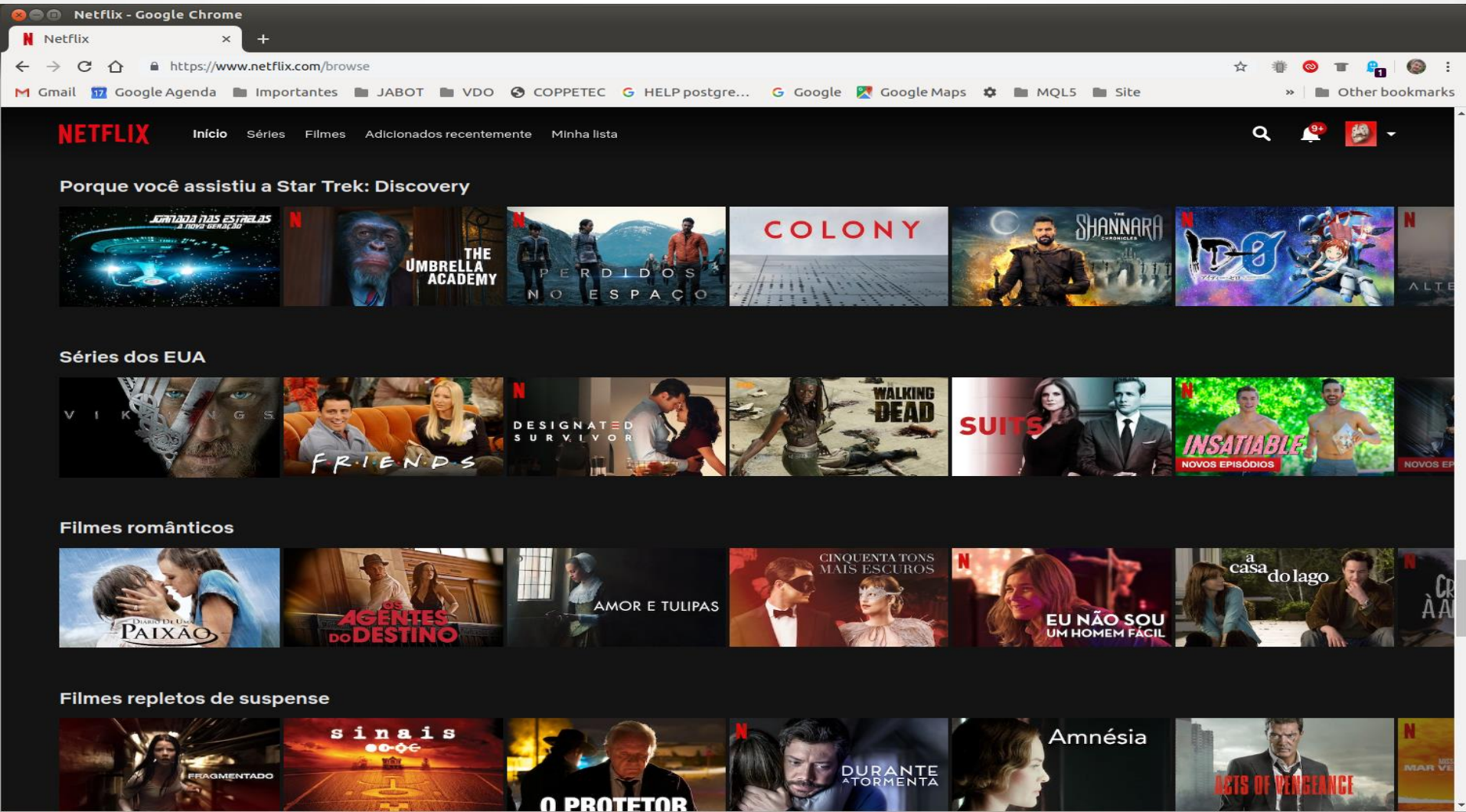


Amazon Basics
USB-C to USB-A Cable



Memory Stick PRO Duo
compatible
micro SD
Adaptor

Netflix - Recomendação



Netflix – recomendação personalizada

Netflix - Google Chrome

Netflix

https://www.netflix.com/title/80232871?bv=80223997&jbp=0&jbr=1

Gmail 17 Google Agenda Importantes JABOT VDO COPPETEC HELP postgre... Google Google Maps MQL5 Site Other bookmarks

NETFLIX Início Séries Filmes Adicionados recentemente Minha lista

Populares na Netflix

FRATURA

CÓPIAS DE VOLTA À VIDA

LAVANDERIA

OUTRA VIDA

VIS VVS NOVOS EPISÓDIOS

ELI

N FILME

FRATURA

99% relevante 2019 14 1h 40min

Sua mulher e a filha desapareceram de um pronto-socorro. Convencido de que o hospital está escondendo algo, ele parte numa busca desesperada para encontrá-las.

▶ ASSISTIR + MINHA LISTA

Estrelando: Sam Worthington, Lily Rabe, Stephen Tobolowsky
Gêneros: Suspenses, Thrillers psicológicos
Cenas e momentos: Sinistros, Suspense no ar

FRATURA

TRAILERS E MAIS

TÍTULOS SEMELHANTES

DETALHES

Em alta

DAYBREAK

dezessete

PRISÃO DE MULHERES

CARA X CARA

ANIMADA

Amazon music

Amazon Music Unlimited - Tenha streaming de mais de 50 milhões de músicas agora. - Google Chrome

Amazon.com.br: Amazon x Prime Music x Amazon Music Unlimited x +

https://music.amazon.com.br/home?ref=dm_ws_lnd_pm_listn_pm_490e581d-5557-4691-a11f-acff8e545064

Gmail 17 Google Agenda Importantes JABOT VDO COPPETEC HELP postgre... Google Google Maps MQL5 Site Other bookmarks

amazon music

Nenhuma música está sendo reproduzida

EXPLORAR

- Início
- Estações
- Playlists

RECENTES

- Reproduzidas
- Adicionadas

BIBLIOTECA

- Álbuns
- Artistas
- Músicas
- Gêneros

PLAYLISTS SEGUIDAS

- Obter Playlists

MINHAS PLAYLISTS

- + Criar Playlist
- Afternoon
- Lara
- Laura
- Ver tudo
- Geraldo Zimbrão

FRESCOR MPB

Novidades

Os Mais Escutados

Playlists recomendadas

Com base no que escutou

VEJA MAIS

NOW That's What I Call Music Vol. 22

Rihanna, Sean Paul, Chamillionaire e muito mais

Best of Avril Lavigne

Avril Lavigne, Gavyn Wright, Butch Walker e muito mais

Eric Clapton and More

Derek & The Dominos, Amboy Dukes Featuring Ted Nugent, B. B. King e muito...

Cyndi Lauper and More

Cyndi Lauper, Anton Fig, Krystal Davis e muito mais

Big Hair Don't Care

Poison, Van Halen, Bon Jovi e muito mais

Classic Rock: British Excellence

The Sweet, Small Faces, Electric Light Orchestra e muito mais

VEJA MAIS

Álbuns recomendados

Com base no que escutou

VEJA MAIS

Tudo gira em torno de você

- Você é o cliente!
- O cliente tem sempre razão!
 - Mas às vezes ele não sabe onde procurar
 - Fica perdido em meio a tantas opções
 - E às vezes ele não sabe nem que quer muito um determinado produto que ele nem conhecia:
 - “Como posso ter passado tanto tempo sem saber que isso existia?”

Cenário Típico

- Existem clientes (usuários)
 - Existem produtos/serviços (itens)
- Queremos ajudar o usuário a escolher o item que ele vai gostar mais!

• Por analogia:

- Quando você entra em uma loja, um vendedor de carne e osso pergunta o que você quer e então lhe recomenda um produto
- Para isso, ele conhece os produtos da loja e obtém alguma informação sobre você:
 - O que você quer? Do que você gosta?



A chave é a informação!

- O vendedor faz uma recomendação!
- Um sistema de recomendação é um “vendedor virtual”
 - Pra isso funcionar, é necessário conhecer os itens, conhecer os usuários e saber o que eles querem.
 - Em algumas situações, sabemos o que o usuário quer: assistir um filme, ouvir uma música. Em outras situações nem isso sabemos.

Sistemas de Recomendação

- Na vida cotidiana, contamos com recomendações de outras pessoas, de boca em boca, cartas de recomendação, resenhas de filmes e livros impressas em jornais ...
- Em um sistema de recomendação típico, as pessoas fornecem recomendações como entradas, que o sistema agrega e direciona aos destinatários apropriados
 - Agregação de recomendações
 - Casa as recomendações com aqueles que procuram recomendações.

Como isso funciona?

- Cada usuário tem um perfil
 - Usuários classificam itens
 - Explicitamente:
 - pontuação de 1 a 5 estrelas
 - Implicitamente:
 - mineração de uso da web
 - Tempo gasto na visualização do item
 - Caminho de navegação
 - Etc.
- O Sistema faz o resto!

Três abordagens básicas

- Recomendação baseada em conteúdo
 - Recomendar itens com base em palavras-chave
 - Semelhante à um problema de recuperação de informações
- Filtragem Colaborativa (CF)
 - Observe o comportamento coletivo dos usuários
 - Observe o histórico do usuário ativo
 - Combine os dois!
- Sistemas Híbridos
 - Combinam a recomendação baseada em conteúdo e a filtragem colaborativa

Descrevendo os itens

- Os itens podem ser descritos através de atributos:
 - Filmes: gênero, diretor, atores, ano, país etc.
 - Músicas: gênero, cantor, autor, ano, língua etc.
 - Roupas: estilo, cor(es), tecidos etc
 - Celulares: memória, câmera, abertura focal, tempo de exposição, número de processadores, slot pra cartão etc...
- Muito subjetivo, nem sempre relevante
- Custoso (muitos itens)

Descrevendo os usuários

- Perfil do usuário:
 - Idade, sexo, estado civil, renda...
 - Mas como obter o gosto do usuário? Um questionário?
 - Muitas perguntas, chato de responder, e às vezes, não reflete o que o usuário quer, apenas o seu humor.
 - Muito subjetivo, nem sempre relevante.
 - Muito custoso (para o usuário)
 - Uma alternativa é perguntar se o usuário gostou ou não de um item, e inferir os gostos do usuário a partir dos atributos desse item.

Recomendação baseada em conteúdo

- Recomenda um item baseado na sua descrição através de atributos e no perfil do usuário
 - Gera as recomendações com base na similaridade dos itens já consumidos pelo usuário.
 - Tanto a descrição do item como o perfil do usuário têm de estar corretos para o resultado ser satisfatório
 - A descrição dos itens é custosa
 - O que são itens semelhantes?
 - Tende a recomendar mais do mesmo! (bolha)

Celulares – bem descritos

tudocelular.com
O site brasileiro de telefonia

Pesquisar

Home Login Fórum > TudoAndroid > TudoWindows > TudoApple > TudoPlanos >

Veja Também Guias e Apps Pr

Comparação em Tamanho Real



LG G8 ThinQ



Samsung Galaxy S10e

Resistência a água	✓	✓
NOTAS		
Custo-benefício	-	6.7 / 10
Hardware	9.1 / 10	8.8 / 10
- Tela	9.4 / 10	8.6 / 10
- Câmera	8 / 10	7.7 / 10
- Desempenho	9.6 / 10	9.4 / 10
REDE		
Sim Card	Nano	Nano
Dual Sim	✓ também disponível single SIM	✓ também disponível single SIM
Gsm	✓ Quad Band (850/900/1800/1900)	✓ Quad Band (850/900/1800/1900)
HSPA+	✓	✓
LTE	✓	✓
Velocidade máxima de download	2000 Mbps	2000 Mbps
Velocidade máxima de upload	318 Mbps	316 Mbps
DADOS TÉCNICOS		
Processador	1x 2.84 GHz Kryo 485 + 3x 2.42 GHz Kryo 485 + 4x 1.8 GHz Kryo 485	4x 1.95 GHz Cortex-A55 + 2x 2.3 GHz Cortex-A75 + 2x 2.7 GHz M4
Chipset	Snapdragon 855 Qualcomm SDM855	SAMSUNG Exynos 9 Octa 9820

Filmes?

Paralelamente, o roteiro escrito por [Scott Silver](#) e o próprio diretor traz um verdadeiro achado, ao estabelecer um subtexto político em torno da transformação de Arthur Fleck no Coringa. Pouco a pouco, a luta de classes chega a Gotham City de forma absolutamente orgânica, com direito a uma referência deliciosa a [Tempos Modernos](#), provocando um levante dos oprimidos junto à elite local, cujo representante maior é... Thomas Wayne. Sim, **Coringa** também passa pela origem do Batman, mais uma vez dialogando com a memória coletiva, entregando uma versão inédita de uma história pra lá de batida.



Claramente inspirado nos filmes urbanos de [Martin Scorsese](#), em especial [Taxi Driver](#) com sua estética das ruas e fotografia suja, **Coringa** ainda apresenta uma apurada direção de arte na construção deste filme de época e um figurino preciso, surrado e ao mesmo tempo recorrente às roupas e cores usuais do personagem-título. Em relação ao restante do elenco, claramente ofuscado por Phoenix, merece destaque a desenvoltura de [Robert De Niro](#) como o apresentador de TV Murray Franklin, óbvia referência (invertida) ao seu papel em [O Rei da Comédia](#), dirigido pelo mesmo Scorsese lá em 1982.

Conteúdo gerado pelo usuário

- O conteúdo gerado pelo usuário tem sido a chave do sucesso para muitas das principais empresas da Web 2.0 de hoje, como Amazon, eBay e Youtube.
- A comunidade agrega valor a esses sites, que, em muitos casos, são quase inteiramente construídos com conteúdo gerado pelo usuário
- Tipos de conteúdo gerado pelo usuário:
 - artigos, avaliações (tripadvisor)
 - vídeos caseiros (youtube)
 - fotos (flickr)
 - avaliações / classificações de itens (todos!)
 - informações coletadas das ações dos usuários on-line (por exemplo, no sistema de recomendação da Amazon).

Trip advisor

Arcos da Lapa - Rio de Janeiro

5.213 avaliações

N.º 155 de 737 atividades em Rio de Janeiro

Pontos turísticos e de interesse, Pontos de interesse, Obras arquitetônicas

📍 Rua Riachuelo 27, Rio de Janeiro, Estado do Rio de Janeiro 20230-010, Brasil

♡ Salvar ➔ Compartilhar

Reserve com antecedência

Bar do Pub Crawl do Rio Hop na Lapa e Ipanema!

EXCURSÕES EM BARES, CLUBES E PUBS

A partir de **R\$ 75,77***

[Mais informações](#)

Excursão completa de um dia pelo Rio de Janeiro

EXCURSÕES EM ÔNIBUS E MINIVAN

A partir de **R\$ 397,77***

[Mais informações](#)

Santa Teresa, Lapa e Cinelândia com Tram Ride e Selarón Passos

EXCURSÕES CULTURAIS

A partir de **R\$ 300,96***

[Mais informações](#)

[Veja mais excursões](#)



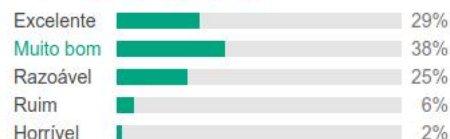
📷 Todas as fotos (2.000)



Opinião dos viajantes

4,0

5.213 avaliações



PREVISÃO DO TEMPO

Fornecido por Weather Underground

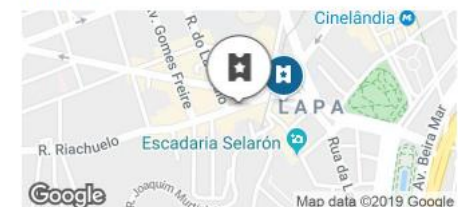
°F | °C

O que dizem os viajantes

“ Ver os **arcos da** cobertura da fundição progresso é magnífico tanto durante o dia quanto a noite, pois além da vista deslumbrante vc se sente mais seguro para contemplar e registrar aquela foto marcante.

“ A **Lapa** é aquele lugarzinho boêmio do Rio de Janeiro, onde você pode encontrar todos os estilos musicais, todo tipo de gente e ainda por cima visitar Santa Terezinha com sua bonde e suas ruas de paralelepídeos.

Contato



📍 Rua Riachuelo 27, Rio de Janeiro, Estado do Rio de Janeiro 20230-010, Brasil

A idéia de filtragem colaborativa

- Tentar prever a opinião que o usuário terá sobre os diferentes itens e poder recomendar os “melhores” itens para cada usuário
- É baseado em:
 - gostos anteriores do usuário
 - opiniões de outros usuários afins
- A FC é um aplicativo típico da Internet
 - deve ser suportado por uma infraestrutura de rede
 - Pelo menos muitos usuários e um servidor
- Não existe aplicativo FC *stand alone*.

Movielens

movielens



311 ☆



top picks



The Shawshank Redemption

★ 5.0



The Usual Suspects

★ 4.9



The Wrong Trousers

★ 5.0



Contact

★ 5.0



As Good as It Gets

★ 5.0



Blue Planet II

★ 5.0



Planet Earth

★ 5.0



One Flew Over the Cuckoo's Nest

★ 4.9

recent releases



Before Someone Gets Hurt

★ 4.1



21 Bridges

★ 3.8



Bait

★ 3.8



Terminator: Dark Fate

★ 3.8



Comrade Drakulich

★ 3.8



Jenny Slate: Stage Fright

★ 4.4



Jojo Rabbit

★ 4.2



Tell Me Who I Am

★ 4.3

Filtragem colaborativa

- Usar a informação gerada pelos usuários
- E se descobrirmos o perfil do usuário através de sua interação com o sistema?
 - Explicitamente: *Likes*, avaliações, notas.
 - Implicitamente: tempo passado em uma página, cliques, ouvir uma música do início ao fim etc.
- E se usarmos as avaliações dos usuários para recomendarmos os itens?
 - Recomendar itens que outros usuários com gostos parecidos ao seu gostaram.

Formalmente

- O sistema possui
 - um conjunto de usuários U
 - um conjunto de itens I
 - uma função de utilidade F , que mede o quão útil é um determinado item i para um determinado usuário u :
 - $F:U \times I \rightarrow R$, onde R é um conjunto totalmente ordenado
 - Para cada usuário $u \in U$ Procura-se itens $i \in I$ que maximizem a função de utilidade.
 - O problema é que F não está definida em todo o espaço $U \times I$
 - É necessário extrapolar seus valores!

Tipos de Filtragem Colaborativa

- Baseadas em Memória
 - trabalham diretamente com valores de interações registradas
 - assumem que não há modelo,
 - são essencialmente baseadas na pesquisa de vizinhos mais próximos
- Baseadas em Modelos
 - assumem um modelo “generativo” subjacente que explica as interações item-usuário
 - tentam descobrir esse modelo para fazer novas previsões

CF baseada em memória

- Dois tipos:
 - baseada na similaridade entre usuários
 - baseada na similaridade entre itens

CF baseada em memória

- Baseada na semelhança entre usuários:
 - Calcula uma medida de similaridade entre o usuário ativo e seus vizinhos
 - Ordena e encontra o usuário ativo (vizinhos)
 - Assume que eles têm o gosto parecido ao usuário ativo
 - Assume que o usuário ativo avalia os itens da mesma forma que seus vizinhos mais próximos
 - Se Fulano, Beltrano e **Sicrano** gostaram dos mesmos filmes que eu, e todos eles gostaram do filme **Coringa**, provavelmente eu vou gostar também.

Os cicloalcanos, também chamados de cicloparafinas ou ciclanos, são hidrocarbonetos cíclicos que possuem pelo menos uma cadeia carbônica fechada.

CF baseada em memória

- Baseada na semelhança entre itens
 - Calcula uma medida de semelhança entre os itens
 - Ordena os itens mais bem avaliados pelo usuário, e encontra os itens mais semelhantes a eles
 - Assume que o usuário irá avaliar estes itens com nota semelhante, já que os itens são semelhantes

CF baseada em modelo

- Usa as informações da interação do usuário para criar uma representação dos usuários e dos itens
- Treina um modelo que aproxima a função de utilidade

Lista de Recomendação

- Calcular a avaliação que o usuário dará a cada um dos itens
- Ordenar da maior para menor avaliação
- Apresentar as n-primeiras avaliações
 - Top-n
- Assume-se que o usuário escolherá os itens que ele mais gosta
 - De qualquer forma, é útil para o usuário ter uma previsão de nota

Mapeamento de usuários e itens

- Um usuário $u_k \in U$ pode ser descrito por atributos
 - Por exemplo, podemos tentar descrever o gosto de um usuário usando valores reais:
 - Terror: 1.5
 - Sci-Fi: 0.7
 - Romântico: 0.1
 - Luta: 1.1
 - $u_k = (1.5; 0.7; 0.1; 1.1) \in \mathbb{R}^4$
- De forma análoga podemos descrever um item

Aproximação da Função de Utilidade

- Se usarmos ***n*** atributos para descrever os usuários e ***m*** atributos para descrever os itens, teremos que construir F' , uma função que aproxima F :
 - $F: U \times I \rightarrow R$
 - $F': R^n \times R^m \rightarrow R'$

Avaliação

- A qualidade da aproximação F' pode ser expressa por sua distância a F :
 - MAE: média de $|F' - F|$
 - RMSE: raiz da média de $(F' - F)^2$

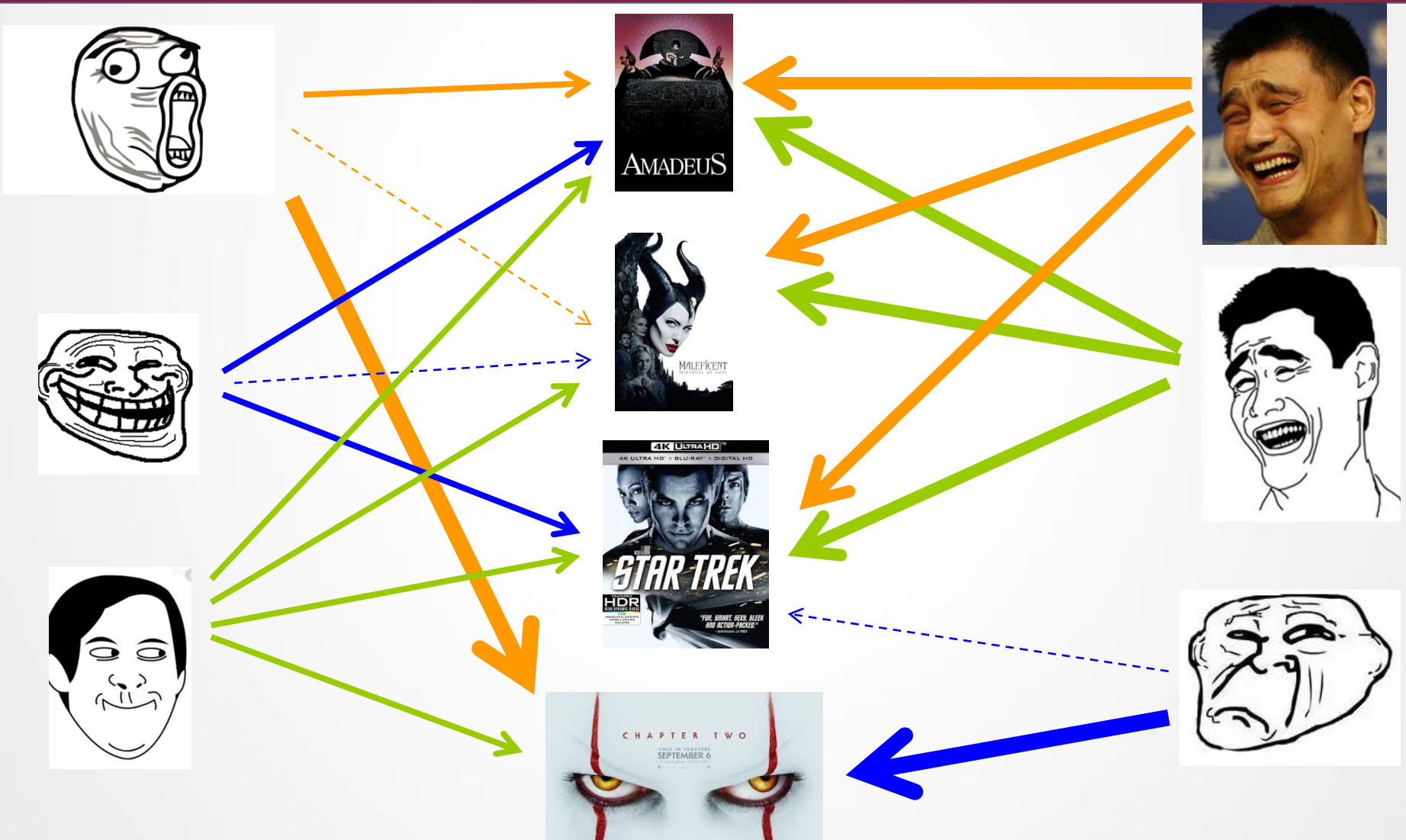
Elementos da Filtragem Colaborativa (1)

- Um sistema de recomendação baseado na Filtragem Colaborativa deve possuir:
 - Um conjunto de Usuários e um conjunto de Itens
 - Cada usuário tem uma lista de itens sobre os quais expressou sua opinião (pode ser um conjunto vazio)
 - Opinião explícita - uma classificação (escala numérica)
 - Opinião implícita - registros de compra ou de uso

Elementos da Filtragem Colaborativa (2)

- Um sistema de recomendação baseado na Filtragem Colaborativa deve possuir:
 - Usuário ativo: para quem a tarefa de previsão de CF é executada
 - Uma métrica para medir a similaridade entre usuários (itens)
 - Um método para selecionar um subconjunto de vizinhos para previsão
 - Um método para prever uma classificação para itens não classificados atualmente pelo usuário ativo.

FC baseada na semelhança entre usuários



FC baseada na semelhança entre usuários

Representação dos dados

Matriz de avaliações usuário x item

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25
a			1		4	5			4		3					2			4		2				
b			4							3							5	1		3					
c		5		4			4						3		5					4		5			
d								3				5				3			4		2			3	
e		3					5			4	5				5					1			5	4	
f			4				1		3	5		4	1		5	4	4		4				3		
g	2	4				4		2			5		1	4	5		4	2	4		5				4
h			2		1		4		3	5		4	2		5	4	5						5		
i		1					3			5				5		4	4		5			4		3	
j			4			4				5			1		5		4		4				4		
k		5				4			2		5		1	5		4		2		4					2
l					3			3				4	1		4		4	2	4					3	
m	5		3					5	3		5	4		5	5	3			4	4	5	4		4	
n			1		4	5				4	5		1	5		4		3		4		4	3		
o			4			4				5		4		5			4	2		5		5		3	
p				4			5								5	4		2	4	4	5	4		2	
q					3			3					1	5		4	4		4			4		3	
r		4			1	4		2					2		5		4				5	4		4	
s			2		4		4			5			1			4		2	4		4		5		
t		1		4			3					4		5	5		4			4				3	
u			2		1		4		3				1		5	4		2	4		5	4			
v					4	5				4	3		5			2					2			5	
w				2			2		3			5				4	5		4	2		3	4		
x	4			5				3		3				4	5					1					
y			1			3				2	3							3	3		5		4		

Passos

1. Para um usuário ativo, o conjunto de suas avaliações é identificado
2. Os usuários mais parecidos com o usuário ativo (de acordo com uma função de similaridade) são identificados (formação de vizinhos)
3. Os itens avaliados por esses usuários semelhantes são identificados
4. Para cada um desses itens, é gerada uma **previsão** da avaliação que seria dada pelo usuário ativo à esse item.
5. Com base nessas classificações previstas, recomenda-se o conjunto dos N itens mais bem avaliados.

Exemplo

	a	b	c	d	e	f	g	h
W	1	1	2	1	2	4	5	4
X	5	?	5	3	2	2	2	1
Y	5	4	4	4	4	2	1	2
Z	4	4	4	4	3	3	3	1

Problemas de Escala

- Um usuário pode interpretar a escala de notas diferentemente do outro
 - Usuários mais “econômicos” x usuários mais “tolerantes”
 - Usuários que gostam de todos os filmes x usuários que não gostam de nenhum filme
- => é preciso normalizar as notas!

Previsão de uma avaliação

- Média simples

$$r_{x,j} = R(u_x, i_j) = u_x + \sum_y (r_{y,j} - u_y)$$

- Média ponderada pela similaridade $\omega_{x,y}$

$$r_{x,j} = R(u_x, i_j) = u_x + \sum_y (r_{y,j} - u_y) \omega_{x,y} / \sum_y \omega_{x,y}$$

- Quantos vizinhos usar?
 - No movielens, entre 5 e 10 produzem um bom resultado

Medidas de Similaridade

- Como avaliar a similaridade de dois usuários?
 - Notas semelhantes para os mesmos itens
 - Quanto mais itens em comum, melhor é a “força” da medida de similaridade
 - E os itens que só foram avaliados por um deles?

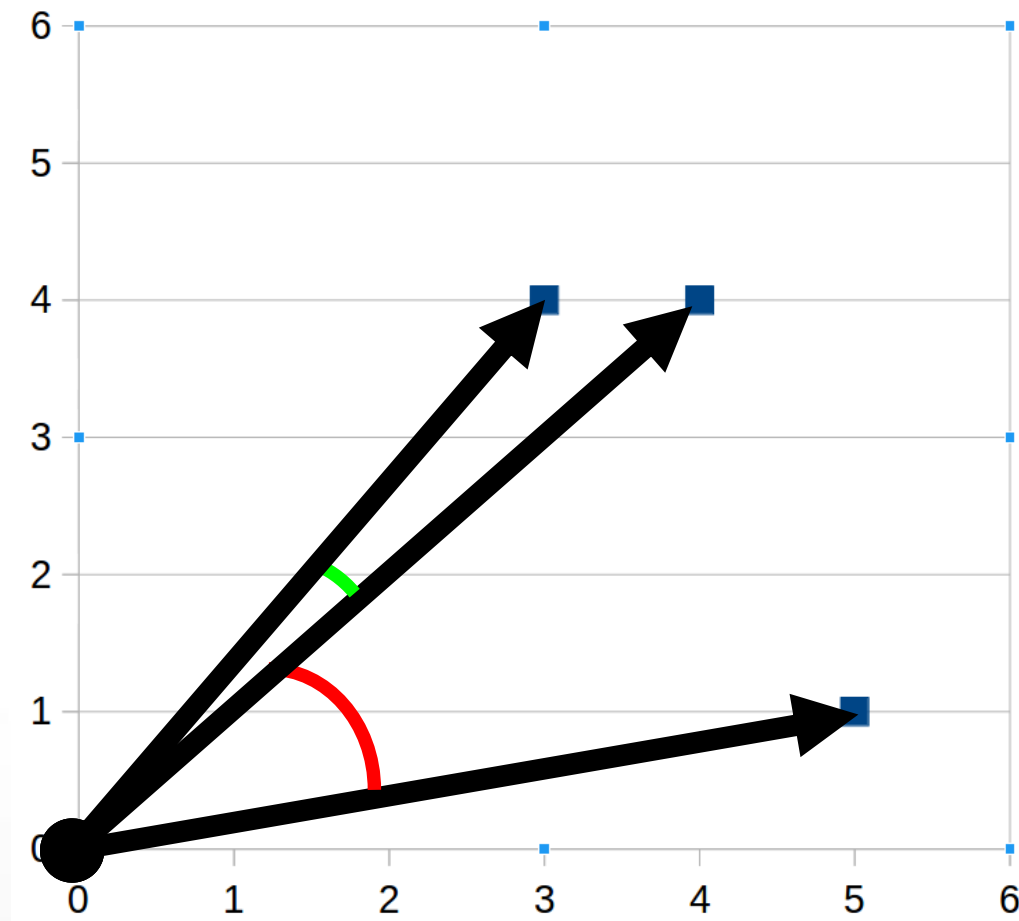
	a	b	c	d	e	f	g	h	i	j	k	l	m	n	o
x	?	?	?	?	4	5	4	4	2	4	5	5	3	2	3
y	3	4	3	5	5	5	4	4	3	5	?	?	?	?	?

Similaridade calculada pelo Cosseno das notas

- Podemos considerar que as notas de um usuário formam um vetor em R^n
- O cosseno entre dois vetores é inversamente proporcional ao ângulo entre eles:
 - Ângulos pequenos indicam que os vetores apontam “na mesma direção

Exemplo

- Usuário 1: (3, 4)
- Usuário 2: (4, 4)
- Usuário 3: (5, 1)



Similaridade de Pearson

- Correlação de Pearson: é uma medida da correlação linear entre duas variáveis X e Y .
- Tem um valor entre $+1$ e -1 ,
 - 1 é correlação linear positiva total,;
 - 0 não é correlação linear; e
 - -1 é correlação linear negativa total.
- Podemos calcular a similaridade entre dois usuários calculando a correlação de Pearson entre suas notas

Fórmula

- A média já está na fórmula
 - não afeta normalizar a matriz de avaliações

$$w_{uv} = \frac{\sum_{j \in I_{uv}} (r_{uj} - r_u)(r_{vj} - r_v)}{\sqrt{\sum_{j \in I_{uv}} (r_{uj} - r_u)^2 \sum_{j \in I_{uv}} (r_{vj} - r_v)^2}}$$

Exemplo na Planilha!

Problemas

- Matriz esparsa
- Não consigo calcular a similaridade entre dois usuários que possuem poucas avaliações em comum
- Muitos itens com poucas avaliações – poucos vizinhos com nota para esse item
- Não consegue prever todas as notas
- Pode deixar de prever justamente itens que teriam notas altas
- Cobertura

CF baseada na semelhança entre itens

- De forma análoga, podemos calcular a similaridade entre itens
- Uma das vantagens é que a semelhança entre itens é pouco afetada por um único usuário
 - Pode ser calculada e armazenada

CF baseada em modelos

- Idéia: Completar a matriz de notas utilizando técnicas de álgebra linear
 - Decomposição em fatores
 - Decomposição SVD
- Idéia: reduzir a dimensionalidade e treinar um classificador
 - Landmarks
 - COFILS

Decomposição em fatores

- Assume que a matriz de avaliações A pode ser expressa como uma multiplicação de matrizes de fatores
 - $A = L \times V^t$
 - Cada fator seria uma variável “latente”, representando um aspecto do gosto do usuário:
 - Terror, musical, suspense...
 - O sistema “descobre” o que são esses atributos

Exemplo

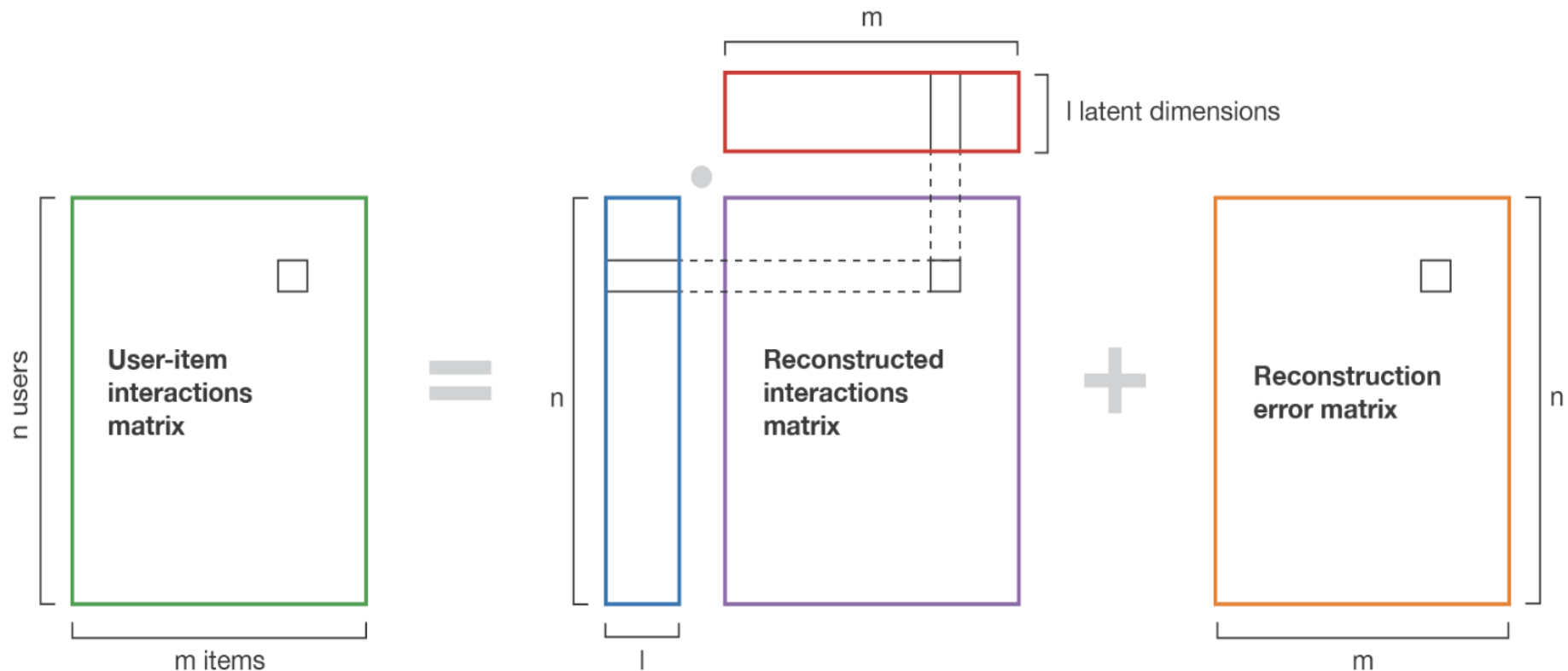
- 4 fatores
- Usuário: (0.6; 1.2; 0.9; 0.7)
- Filme: (2.1; 0.5; 0.7; 1.0)
- Avaliação:

$$2.1 \times 0.6 + 1.2 \times 0.5 + 0.9 \times 0.7 + 0.7 \times 1.0$$

$$\Rightarrow \mathbf{3.19}$$

- Consigo calcular todas as notas!

Fatores



The **user-item interactions matrix** is assumed to be equal to...

... the **dot product** of a **user matrix** and a **transposed item matrix**...

... plus some **reconstruction error**

$$M \approx XY^T$$

Decomposição SVD

- Uma matriz A pode ser expressa como produto de três matrizes $A = USV^t$, onde S é uma matriz diagonal.
- Os valores de U_k e I_k são os atributos de cada usuário e cada filme
- O produto entre eles é usado para calcular a nota

Landmarks

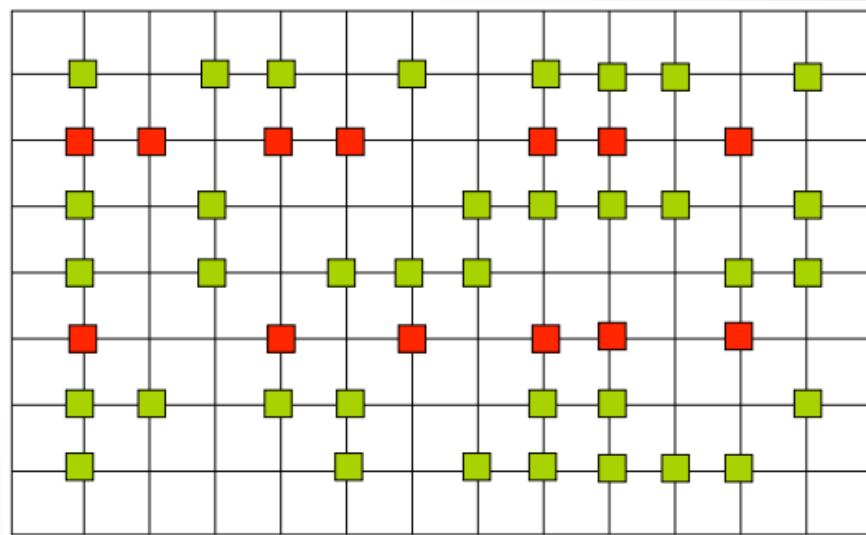
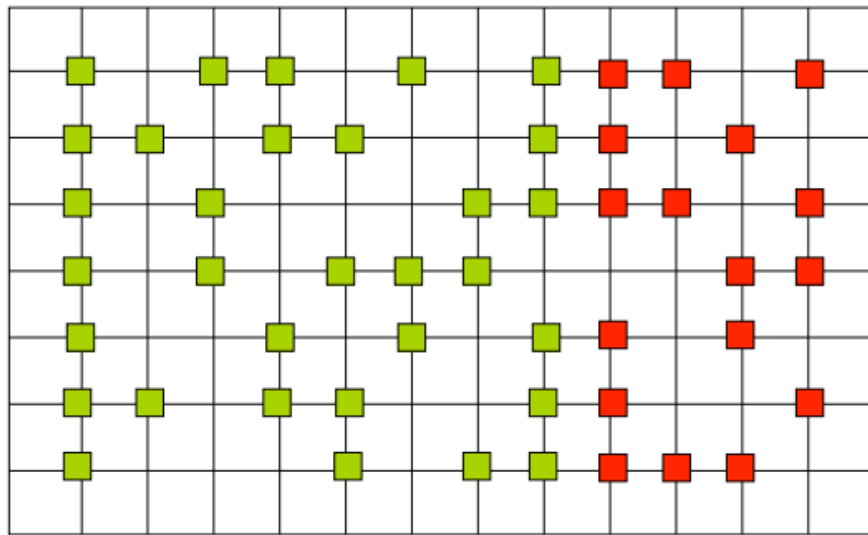
- Tese de doutorado
- Idéia: escolher K usuários com muitas avaliações (landmarks)
 - Landmarks podem ser explicados como **gurus**.
- Calcular a similaridade entre cada usuário do sistema e os landmarks.
 - Isso produz um vetor de K dimensões para cada usuário
 - Agora, calculamos a similaridade entre cada usuário usando esses vetores
 - Vantagem: consigo calcular a similaridade entre todos os usuários
 - Se um item possui nota, consigo fazer a previsão
 - O mesmo pode ser aplicado a itens

COFILS

- Tese de doutorado
- Aplicar uma redução de dimensionalidade
 - Landmarks
 - SVD
 - PCA
- Treinar uma rede neural (ou qualquer outro preditor/classificador) recebendo como entrada o usuário, o item e a nota nesse novo espaço R^n

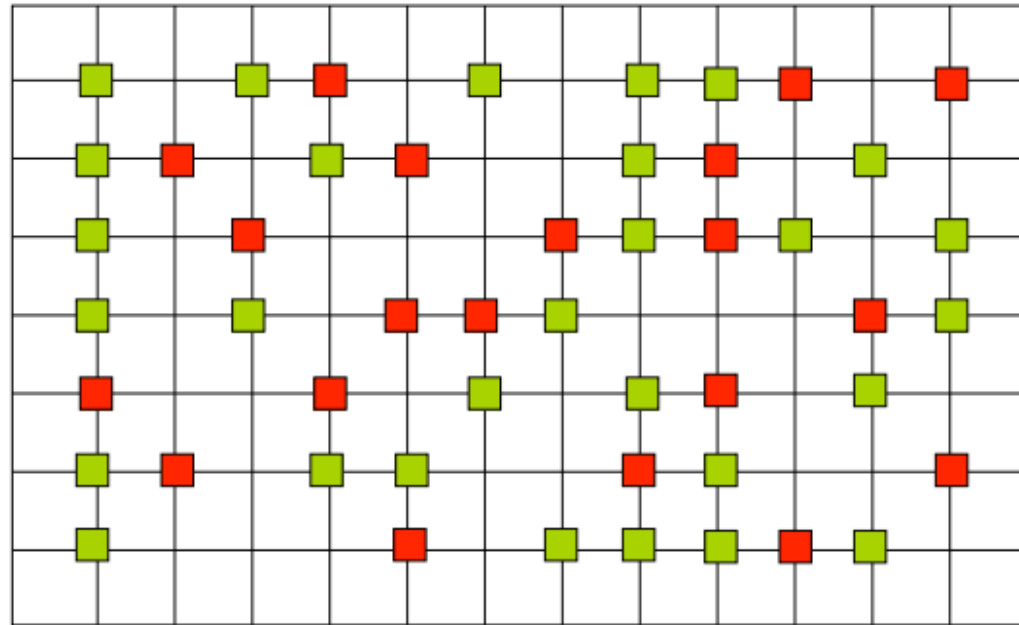
Cuidados ao realizar os experimentos

- Matriz de avaliações: não podemos separar em teste e treino usando colunas e linhas



Selecionar aleatoriamente notas

- Dados: conjunto de avaliações que os usuários fazem dos itens
 - Selecionar aleatoriamente avaliações de todos os usuários e itens para o conjunto de teste e de treinamento
 - k-fold é uma boa abordagem



Cold Start

- Como recomendar itens para os usuários que não avaliaram nada?
 - Quais os melhores itens para oferecer ao usuário para avaliar?
- Como recomendar produtos novos, que ainda não foram avaliados?
 - Incentivar alguns usuários a avaliá-los?
 - Quais usuários teriam maior impacto?

Obrigado!

Questões!