



Sistemas de Recomendação

Camila Tormena
João Guilherme Packer
Marcos Eduardo Crivellaro

Agenda

- Introdução
 - Objetivo
 - Definição e histórico
 - Motivações e Aplicações
 - Métricas

Agenda

- Algoritmos
 - Técnicas Não-Personalizadas
 - Como Não ordenar pela média
 - Como ordenar pela média
 - Filtragem Colaborativa
 - Baseada em Vizinhos
 - Filtragem Baseada em Conteúdo
 - Métodos Híbridos
 - Métodos Utilizados pelo Netflix

Agenda

- Implementação

- Filtragem colaborativa baseada em usuários
 - Filtragem colaborativa baseada em itens
 - Ferramentas

- Estudos de caso

- Netflix
 - Google AdSense

Agenda

- Desafios e limitações
 - Esparcidade
 - The Cold-Start Problem
 - Fraude
 - Filter Bubble
- Considerações Finais

Seção 1

Introdução

Objetivo

- Visão geral dos sistemas de recomendação
- Abordagens e algoritmos
- Diferentes usos
- Benefícios
- Exemplos reais

Definição e histórico

- “O objetivo dos sistemas de recomendação (SR) é gerar recomendações válidas para um conjunto de usuários, de itens que possam interessá-los” (MELVILLE; SINDHWANI, 2010)
- Personalizadas x Não personalizadas
- “Obter recomendações de fontes confiáveis é um componente importante do processo natural de tomada de decisões” (MELVILLE; SINDHWANI, 2010).
- Consumismo + Internet

Definição e histórico

- Problema com sobrecarga de informações
 - Mais opções nem sempre é melhor
- As pessoas geralmente apoiam-se em recomendações de outras pessoas para suas decisões diárias
- Primeiro SR comercial: Tapestry
- Informações explícitas x implícitas

Definição e histórico

- Abordagens
 - Filtragem colaborativa
 - Filtragem baseada em conteúdo
 - Métodos híbridos
- Área de pesquisa independente surgiu nos anos 90
- Interesse crescente
 - Papel importante em sites altamente conceituados
 - Conferências e workshops dedicados
 - Sessões em conferências de outras áreas
 - Cursos dedicados à área
 - Edições especiais em revistas acadêmicas

Definição e histórico

1988

A rule-based
message filtering
system
(POLLOCK,
1988)

1992

Tapestry
(GOLDBERG
et al., 1992)

1990

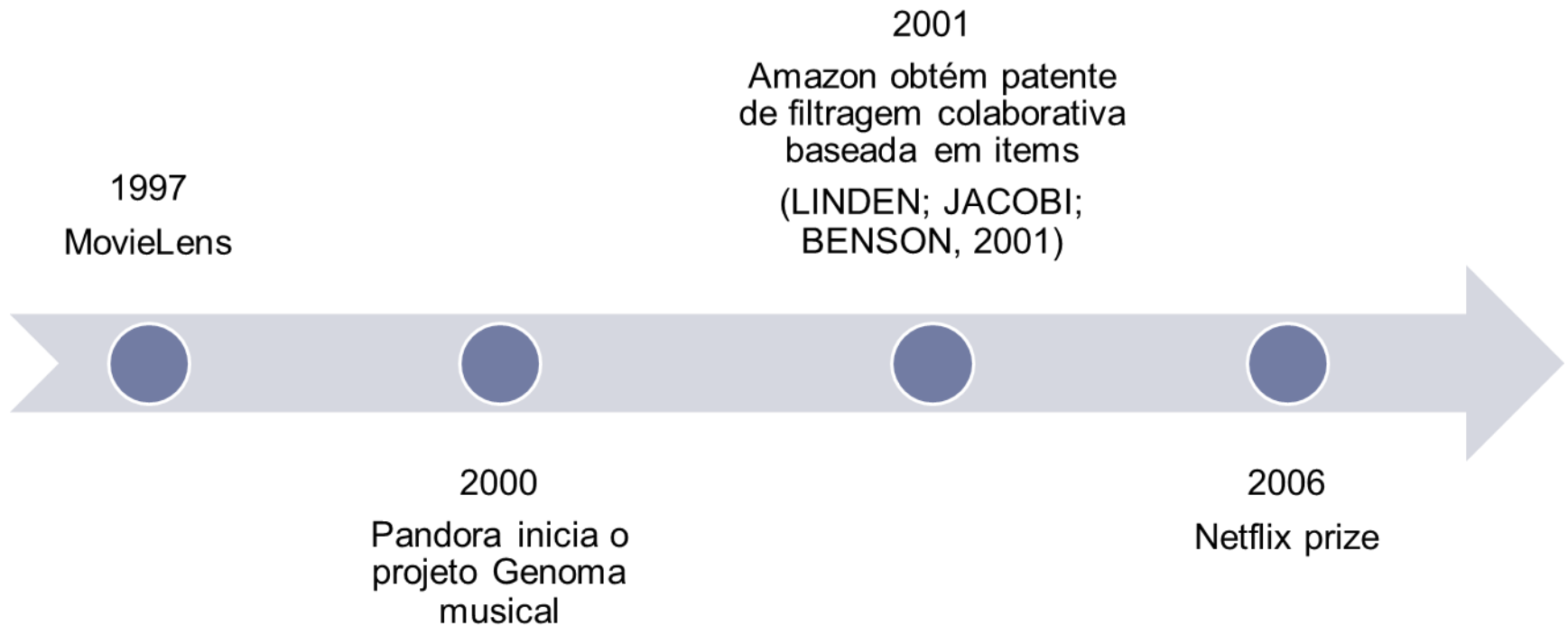
MAFIA: An active mail-
filter agent for an
intelligent document
processing support
(LUTZ; KLEIST-RETZOW;
HOERNING, 1990)

1994

GroupLens
(RESNICK et
al., 1994)



Definição e histórico



Motivações e aplicações

- “Os SR provaram ser uma ótima abordagem para lidar com o problema de sobrecarga de informações citado” (RICCI; ROKACH; SHAPIRA, 2011)
- “1/3 dos consumidores que notam as recomendações acabam comprando algo baseado nelas” (GROSSMAN, 2013)
- SRs são sistemas complexos
- Informações sobre itens x usuários

Motivações e aplicações

- Cinematch
 - Sistema de recomendações do Netflix
- Prêmio Netflix oferecido em 2006
 - 1 milhão de dólares para uma melhora de 10%
 - “Uma melhora de 10% traria para a empresa lucros além do um milhão de dólares oferecido” (THOMPSON, 2008)
 - Ganhador: BellKor's Pragmatic Chaos (2009)
- Nem sempre as características em comum dos filmes são identificáveis

Motivações e aplicações

- Motivos para implantar um SR
 - Aumentar o número de itens vendidos
 - Vender itens mais diversificados
 - Aumentar a satisfação dos usuários
 - Aumentar a fidelidade dos usuários
 - Melhorar o gerenciamento dos itens
- SRs estão sujeitos à falhas
 - Walmart acusado de racismo
- Problema está na consistência dos dados!

Métricas

- Identificar o melhor algoritmo de recomendação é um desafio
 - Discordância sobre os atributos e métricas
- Problemas ao avaliar os algoritmos:
 - Algoritmos dependem do conjunto de dados
 - Objetivos da avaliação podem variar

Métricas

- A maioria dos métodos foca na precisão
- “A falta de padronização leva a uma grande quantidade de métricas sendo utilizadas, tornando-se difícil comparar os resultados de diferentes publicações” (HERLOCKER et al., 2004)
- As métricas de avaliação podem ser divididas em quatro grandes classes:
 - Precisão preditiva
 - Precisão de classificação
 - Precisão de ranking
 - Não precisão

Métricas

- Precisão preditiva (Predictive accuracy)
 - Avaliações estimadas x avaliações do usuário
 - Adequada nos cenários em que uma predição correta das avaliações para todos os itens é de alta importância
 - Root mean squared error (RMSE) foi a métrica utilizada na competição do Netflix
 - Fáceis de computar e compreender

Métricas

- Precisão de classificação (Classification accuracy)
 - Mede a qtd de classificações corretas e incorretas
 - Adequada para aplicações levar os usuários a tomar certas ações como comprar produtos ou serviços
- Precisão de ranking (Rank accuracy)
 - Verifica a habilidade de estimar a ordem correta dos itens em relação à preferência do usuário
 - Adequada para sistemas que oferecem ao usuário uma longa lista ordenada de itens recomendados.

Métricas

- Além da precisão
 - Sistemas precisam também convencer usuários
 - Depende de características individuais
 - “Um SR deve inspirar confiança, possuir uma lógica razoavelmente transparente, apresentar itens novos, prover detalhes sobre os itens, incluir fotos e avaliações da comunidade, e apresentar maneiras de refinar as recomendações” (SWEARINGEN; SINHA, 2001)

Seção 2

Algoritmos

Algoritmos de Sistemas de Recomendação

- Utilizam como base informações e atributos de usuários e itens para recomendar que estão disponíveis no sistema. Estes são os principais componentes avaliados com base em diferentes critérios, cada qual com sua



Categorias

- Técnicas Não-Personalizadas
 - Como NÃO ordenar pela média
 - Como ordenar pela média
- Filtragem Colaborativa
 - Baseada em Vizinhos
- Filtragem Baseada em Conteúdo
- Métodos Híbridos

Como Não ordenar pela média

- Nota do item =
(Avaliações positivas) - (Avaliações negativas)

twerkapnoia
 Twerkathon
 Twerkative
 Twerk attack
 Twerkbortion
 Twerk burn
 twerk butter
 Twerk Cluster
 Twerked
 twerkee
 Twerken
 Twerker
 Twerker's Back
 Twerkers' compensation
 Twerk Ethic
 Twerkey

3. Twerk

The vigorously shaking of your Gluteus Maximus

That ratchet knows how to twerk.

mark as favorite buy twerk mugs & shirts

booty twerk gluteus ass ratchet.

by A34 Feb 7, 2013 add a video

2831 up, 1799 down



4. twerk

Ghetto dancing

"when i get hammered tonight, i'm twerking my heart out"

mark as favorite buy twerk mugs & shirts

black girl ass alcohol ghetto ratchet

by GabzS456 Dec 31, 2012 add a video

2820 up, 2191 down



Como Não ordenar pela média

- Nota do item = média das avaliações = $(\text{Avaliações positivas}) / (\text{Total de avaliações})$



See Color Options

TY Beanie Boos - Waddles - Penguin

~~\$9.99~~ **\$7.44** ✓ Prime

Order in the next **23 minutes** and get it by Wednesday, Oct 9.
Eligible for FREE Super Saver Shipping.

More Buying Choices

\$2.29 new (40 offers)

★★★★★ (294)

Show only TY Beanie Boos items



Baby Einstein Take Along Tunes

~~\$9.99~~ **\$8.40** ✓ Prime

Order in the next **23 minutes** and get it by Wednesday, Oct 9.
Eligible for FREE Super Saver Shipping. [See Details](#)

More Buying Choices

\$4.71 new (68 offers)

★★★★★ (1,943)

Show only Baby Einstein items

Como ordenar pela média

- Nota do item = Limite inferior do intervalo de confiança da pontuação Wilson para um parâmetro Bernoulli

$$\left(\hat{p} + \frac{z_{\alpha/2}^2}{2n} \pm z_{\alpha/2} \sqrt{[\hat{p}(1 - \hat{p}) + z_{\alpha/2}^2/4n]/n} \right) / (1 + z_{\alpha/2}^2/n).$$

Filtragem Colaborativa

- Algoritmo da filtragem baseada em vizinhos
 1. Atribua um peso para todos os usuários de acordo com a similaridade com o usuário atual;
 2. Selecione k usuários que tenham a maior similaridade possível com o usuário atual - que normalmente é chamado de vizinhança;
 3. Compute uma predição de uma combinação ponderada das avaliações dos vizinhos selecionados.

Exemplo de Algoritmo

- Correlação do Coeficiente de Pearson

$$s_{a,u} = \frac{\sum_{i=1}^m [(r_{a,i} - \bar{r}_a)(r_{u,i} - \bar{r}_u)]}{\sqrt{\sum_{i=1}^m (r_{a,i} - \bar{r}_a)^2 \sum_{i=1}^m (r_{u,i} - \bar{r}_u)^2}}$$

Exemplo de Algoritmo

- Alternativa para a correlação de Pearson (passo 1)

$$\begin{aligned}w_{a,u} &= \cos(\vec{r}_a, \vec{r}_u) = \frac{\vec{r}_a \cdot \vec{r}_u}{\|\vec{r}_a\|_2 \times \|\vec{r}_u\|_2} \\&= \frac{\sum_{i=1}^m r_{a,i} r_{u,i}}{\sqrt{\sum_{i=1}^m r_{a,i}^2} \sqrt{\sum_{i=1}^m r_{u,i}^2}}\end{aligned}$$

Exemplo de Algoritmo

- No passo 3, as previsões geralmente são computadas como a média ponderada dos desvios da média do vizinho.

$$p_{a,i} = \bar{r}_a + \frac{\sum_{u=1}^n [(r_{u,i} - \bar{r}_u) w_{a,u}]}{\sum_{u=1}^n w_{a,u}}$$

Filtragem Colaborativa

- Métodos baseados no usuário:
 - Ao invés de recomendar itens semelhantes aos visualizados anteriormente pelo usuário, recomendar itens semelhantes aos de interesses em comum de outros usuários;
- Alguns métodos que seguem essa abordagem:
 - Cosseno;
 - Pearson;
 - Vizinhos mais próximos;
 - etc;

Filtragem Colaborativa

- A abordagem mais comum é utilizar uma matriz que lista usuários e itens;
- Necessário que os usuários avaliem os itens.

Filtragem Colaborativa

- Características Positivas:
 - Produz recomendações “inesperadas”;
 - Não depende de preferências anteriores do usuário;
 - Produz recomendações de alta qualidade: qualidade de um item avaliada pelos próprios usuários (público alvo).
- Características Negativas:
 - Avaliações Esparsas;
 - Primeiras Avaliações;
 - Cold-Start.

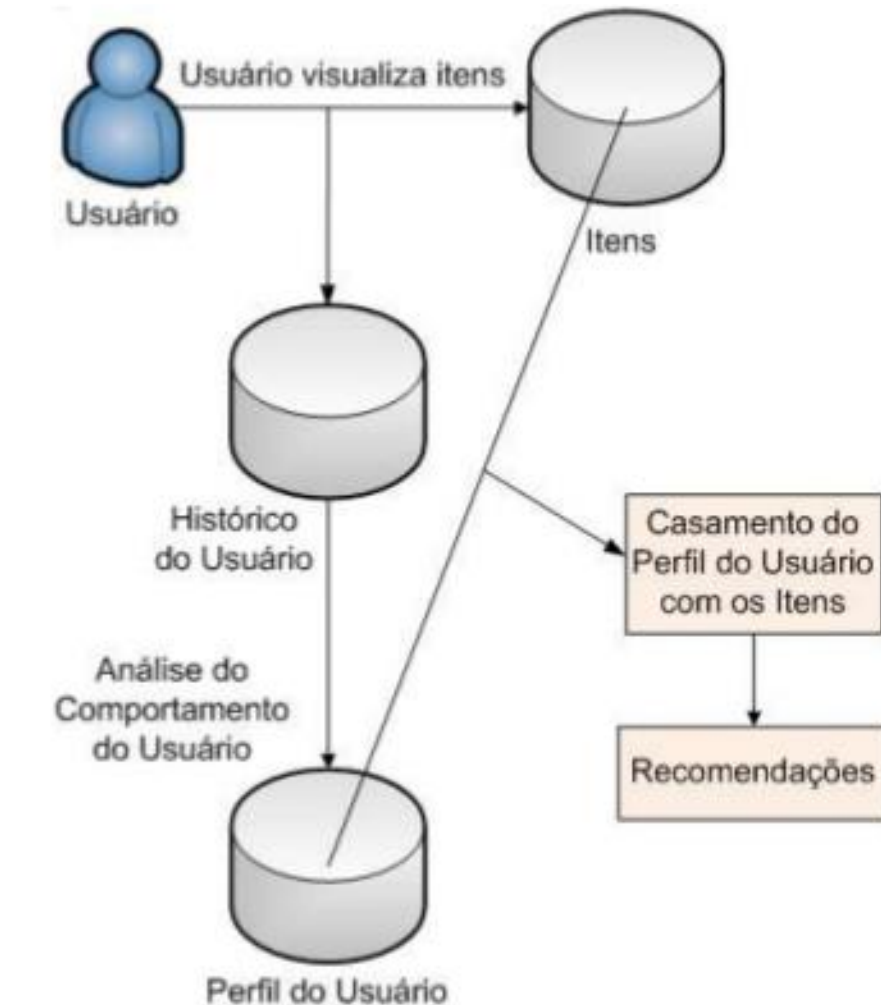
Filtragem Baseada em Conteúdo

- Método baseado em recomendação de itens semelhantes aos que determinado usuário selecionou anteriormente.
- A filtragem de informação é baseada na análise do conteúdo do item (descrito por suas características ou atributos) e no perfil do usuário (que mantém registro de itens selecionados anteriormente);
- Conceito: usuários tem tendência a se interessar por itens semelhantes aos que já comprou anteriormente;

Filtragem Baseada em Conteúdo

- Meios de determinar similaridade e recomendações relevantes:
 - Índices de busca booleana;
 - Filtragem probabilística;
 - Modelos vetoriais;
 - Etc.
- Avaliação do item:
 - Usuário informa sua opinião explicitamente ou o sistema registra itens adquiridos ou visitados pelo usuário.

Filtragem Baseada em Conteúdo



Filtragem Baseada em Conteúdo

- Características Positivas:
 - Usuário Independente;
 - Novo Item.
- Características Negativas:
 - Análise de Conteúdo Limitada;
 - Especialização;
 - Novo Usuário.

Filtragem Baseada em Conteúdo

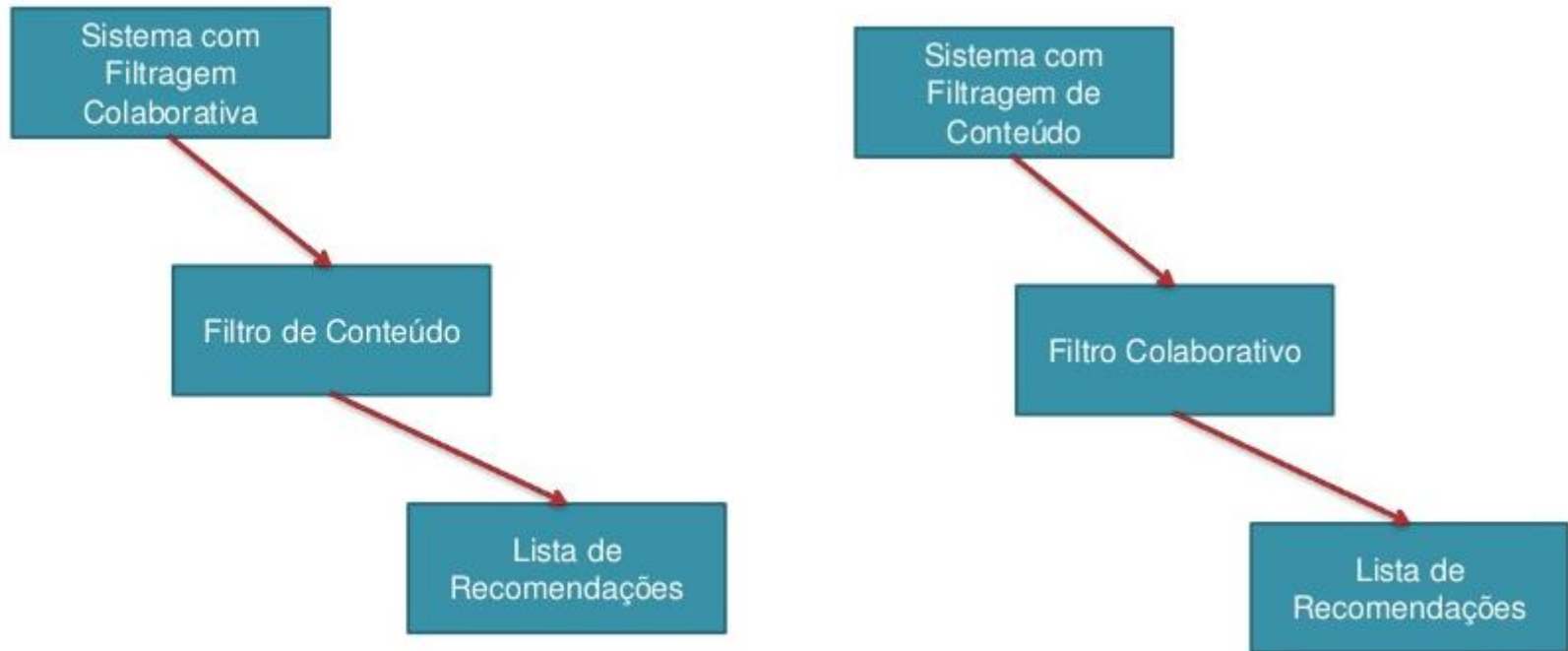
- Muito da pesquisa desta área tem focado em recomendar itens com conteúdo textual associado
 - páginas web, livros e filmes; onde o conteúdo associado como descrições e avaliações de usuários estão disponíveis.
- Portanto, várias abordagens tratam este problema como uma tarefa de recuperação de informação, onde o conteúdo associado com as preferências do usuário é tratado como uma consulta;
- Documentos não-avaliados recebem uma nota de relevância/similaridade de acordo com esta consulta.

Filtragem Baseada em Conteúdo

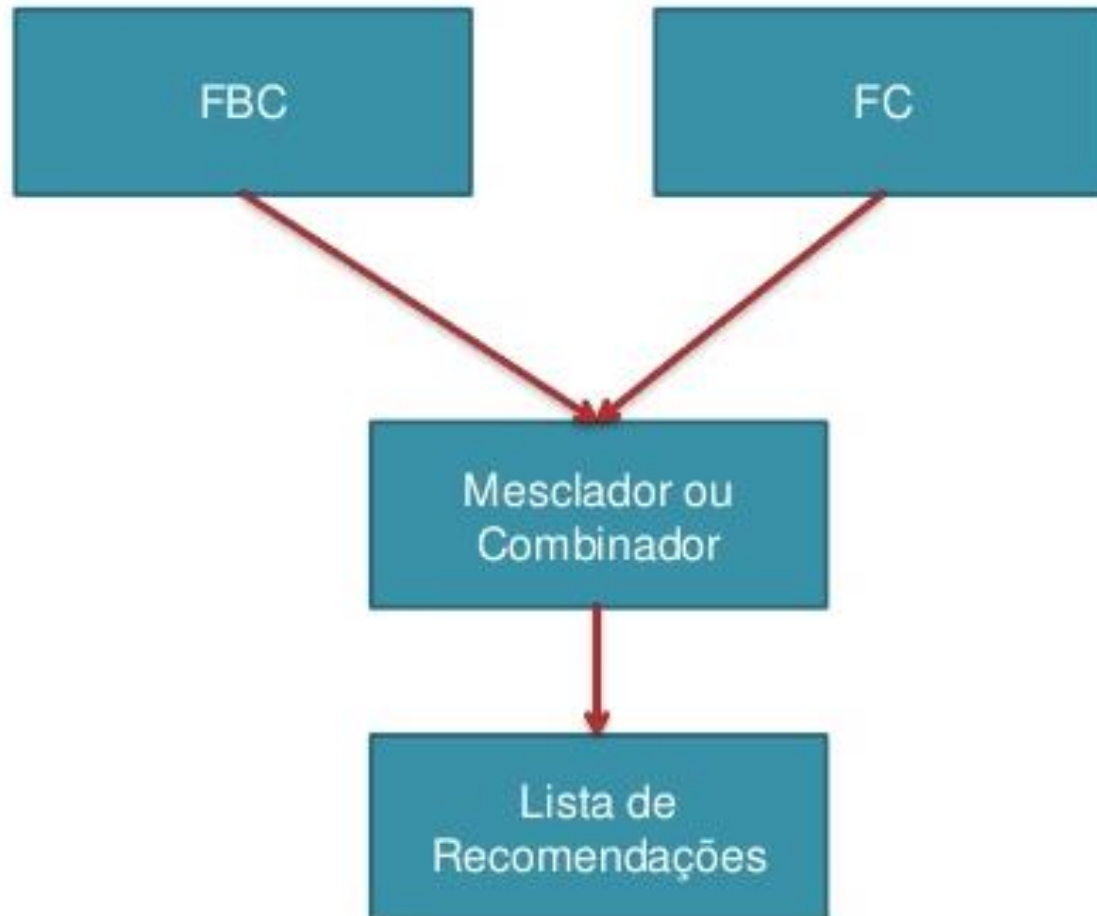
- Uma alternativa para abordagens de recuperação de informação é tratar recomendações como uma tarefa de classificação:
 - Cada exemplo representa o conteúdo de um item, e as avaliações anteriores do usuário são usadas como rótulos para estes.
- Algoritmos utilizados nesta classificação:
 - Vizinho K-mais-próximo;
 - Árvores de Decisão;
 - Classificador de Bayes;
 - Redes Neurais.

Métodos Híbridos

- A filtragem de informação é baseada na análise do conteúdo do item e no perfil do usuário. As configurações possíveis são:



Métodos Híbridos



Métodos Utilizados pelo Netflix

- Matrix factorization (ou também matrix decomposition):
 - Método híbrido;
 - Em sua forma base, caracteriza ambos itens e usuários por vetores;
 - Há vários métodos de diferentes de aplicar, cada qual destinado a determinadas classes de problemas;
 - Ex: na análise numérica, para equações lineares, pode ser utilizada decomposição LU (lower upper).
 - Consiste da fatora  o de uma matriz em produtos de matrizes.

Métodos Utilizados pelo Netflix

- Restricted Boltzmann Machines (RBM):
 - Base em filtragem colaborativa;
 - Aplica distribuição de probabilidades em seu conjunto de parâmetros de entrada, de acordo com critérios específicos.
 - Tipo padrão de RBM compõe:
 - Unidades visíveis (v) e escondidas (h) de valores booleanos e pesos de compensação (a para “ v ” e b para “ h ”);
 - Uma matriz simétrica de pesos W associada as unidades visíveis e escondida;
 - Energia dada por:

$$E(v, h) = - \sum_i a_i v_i - \sum_j b_j h_j - \sum_i \sum_j h_j w_{i,j} v_i$$

Métodos Utilizados pelo Netflix

- Matrix factorization;
- Restricted Boltzmann Machines;
- Linear regression;
- Logistic regression;
- Elastic nets;
- Singular Value Decomposition;
- Markov Chains
- Latent Dirichlet Allocation
- Association Rules
- Gradient Boosted Decision Trees
- Random Forests
- Técnicas de clustering (ex: Affinity Propagation).

Seção 3

Caso de uso

Implementação de um sistema de recomendação simples

Implementação de um SR

- Implementação manual de um algoritmo simples para realizar recomendações
 - Filtragem colaborativa baseada em usuários
 - Filtragem colaborativa baseada em itens
- Utilização de ferramentas para implementação

Coletando preferências

● Representando informações

```
# A dictionary of movie critics and their ratings of a small
# set of movies
critics={'Lisa Rose': {'Lady in the Water': 2.5, 'Snakes on a Plane': 3.5,
    'Just My Luck': 3.0, 'Superman Returns': 3.5, 'You, Me and Dupree': 2.5,
    'The Night Listener': 3.0},
    'Gene Seymour': {'Lady in the Water': 3.0, 'Snakes on a Plane': 3.5,
    'Just My Luck': 1.5, 'Superman Returns': 5.0, 'The Night Listener': 3.0,
    'You, Me and Dupree': 3.5},
    'Michael Phillips': {'Lady in the Water': 2.5, 'Snakes on a Plane': 3.0,
    'Superman Returns': 3.5, 'The Night Listener': 4.0},
    'Claudia Puig': {'Snakes on a Plane': 3.5, 'Just My Luck': 3.0,
    'The Night Listener': 4.5, 'Superman Returns': 4.0,
    'You, Me and Dupree': 2.5},
    'Mick LaSalle': {'Lady in the Water': 3.0, 'Snakes on a Plane': 4.0,
    'Just My Luck': 2.0, 'Superman Returns': 3.0, 'The Night Listener': 3.0,
    'You, Me and Dupree': 2.0},
    'Jack Matthews': {'Lady in the Water': 3.0, 'Snakes on a Plane': 4.0,
    'The Night Listener': 3.0, 'Superman Returns': 5.0, 'You, Me and Dupree': 3.5},
    'Toby': {'Snakes on a Plane': 4.5, 'You, Me and Dupree': 1.0, 'Superman Returns': 4.0}}
```

Encontrando usuários similares

● Identificação de similaridade através da Distância euclidiana

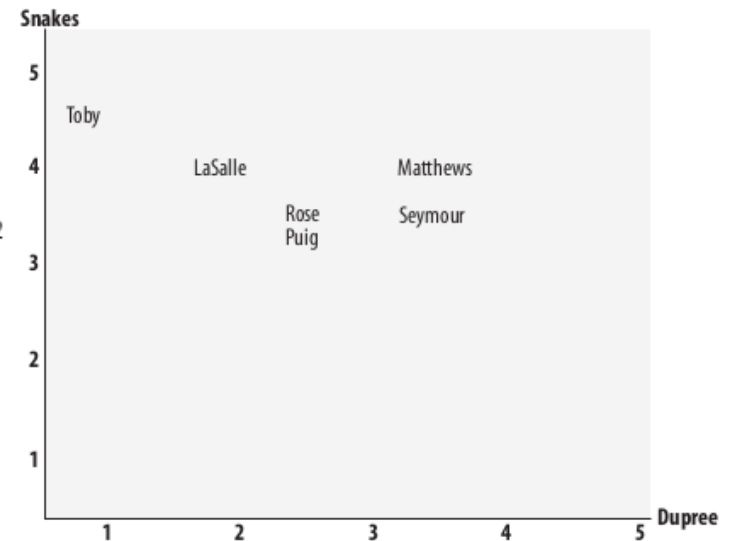
```
from math import sqrt

# Returns a distance-based similarity score for person1 and person2
def sim_distance(prefs, person1, person2):
    # Get the list of shared_items
    si={}
    for item in prefs[person1]:
        if item in prefs[person2]:
            si[item]=1

    # if they have no ratings in common, return 0
    if len(si)==0: return 0

    # Add up the squares of all the differences
    sum_of_squares=sum([pow(prefs[person1][item]-prefs[person2][item],2)
                        for item in prefs[person1] if item in prefs[person2]])

    return 1/(1+sum_of_squares)
```



Classificando usuários

- É verificada a compatibilidade pela comparação de um usuário com todos os outros, através da métrica de distância previamente definida
- Exemplo: Distância Euclidiana

```
# Returns the best matches for person from the prefs dictionary.  
# Number of results and similarity function are optional params.  
def topMatches(prefs, person, n=5, similarity=sim_pearson):  
    scores=[(similarity(prefs, person, other), other)  
            for other in prefs if other!=person]  
  
    # Sort the list so the highest scores appear at the top  
    scores.sort()  
    scores.reverse()  
    return scores[0:n]
```

Recomendando itens

- Problemas ao considerar um único usuário similar
 - Exclui filmes que o mesmo não assistiu e inclui filmes que apenas ele gostou
- Solução: pontuação ponderada para os itens

Critic	Similarity	Night	S.xNight	Lady	S.xLady	Luck	S.xLuck
Rose	0.99	3.0	2.97	2.5	2.48	3.0	2.97
Seymour	0.38	3.0	1.14	3.0	1.14	1.5	0.57
Puig	0.89	4.5	4.02			3.0	2.68
LaSalle	0.92	3.0	2.77	3.0	2.77	2.0	1.85
Matthews	0.66	3.0	1.99	3.0	1.99		
Total			12.89		8.38		8.07
Sim. Sum			3.84		2.95		3.18
Total/Sim. Sum			3.35		2.83		2.53

Relacionando itens

- Utilizado em sites de compras on-line
 - Não possui muitas informações sobre o usuário
- É possível utilizar os métodos implementados anteriormente
 - Basta inverter o dicionário e chamar a função de similaridade:

```
{ 'Lisa Rose': { 'Lady in the Water': 2.5, 'Snakes on a Plane': 3.5 },  
  'Gene Seymour': { 'Lady in the Water': 3.0, 'Snakes on a Plane': 3.5 }}
```

```
{ 'Lady in the Water': { 'Lisa Rose': 2.5, 'Gene Seymour': 3.0 },  
  'Snakes on a Plane': { 'Lisa Rose': 3.5, 'Gene Seymour': 3.5 }} etc..
```

```
>> reload(recommendations)  
>> movies=recommendations.transformPrefs(recommendations.critics)  
>> recommendations.topMatches(movies, 'Superman Returns')  
[(0.657, 'You, Me and Dupree'), (0.487, 'Lady in the Water'), (0.111, 'Snakes on a  
Plane'), (-0.179, 'The Night Listener'), (-0.422, 'Just My Luck')]
```

Filtragem baseada em itens

- Computar recomendações a partir todas as avaliações envolve muitas operações
 - Depende do número de usuários e de avaliações de cada usuário
- Filtragem colaborativa baseada em itens é mais adequada para grandes volumes de dados
 - Cálculos realizados antecipadamente
 - Similaridade entre itens muda com menos frequência

Construindo o conjunto de similaridades

- Etapas:
 - Transformar o dicionário para o formato itens => classificações
 - Iterar sobre cada item e calcular o valor de similaridade com os outros
 - Criar um dicionário com a lista dos itens mais similares
- Precisa ser executado apenas para manter a lista de similaridades atualizada

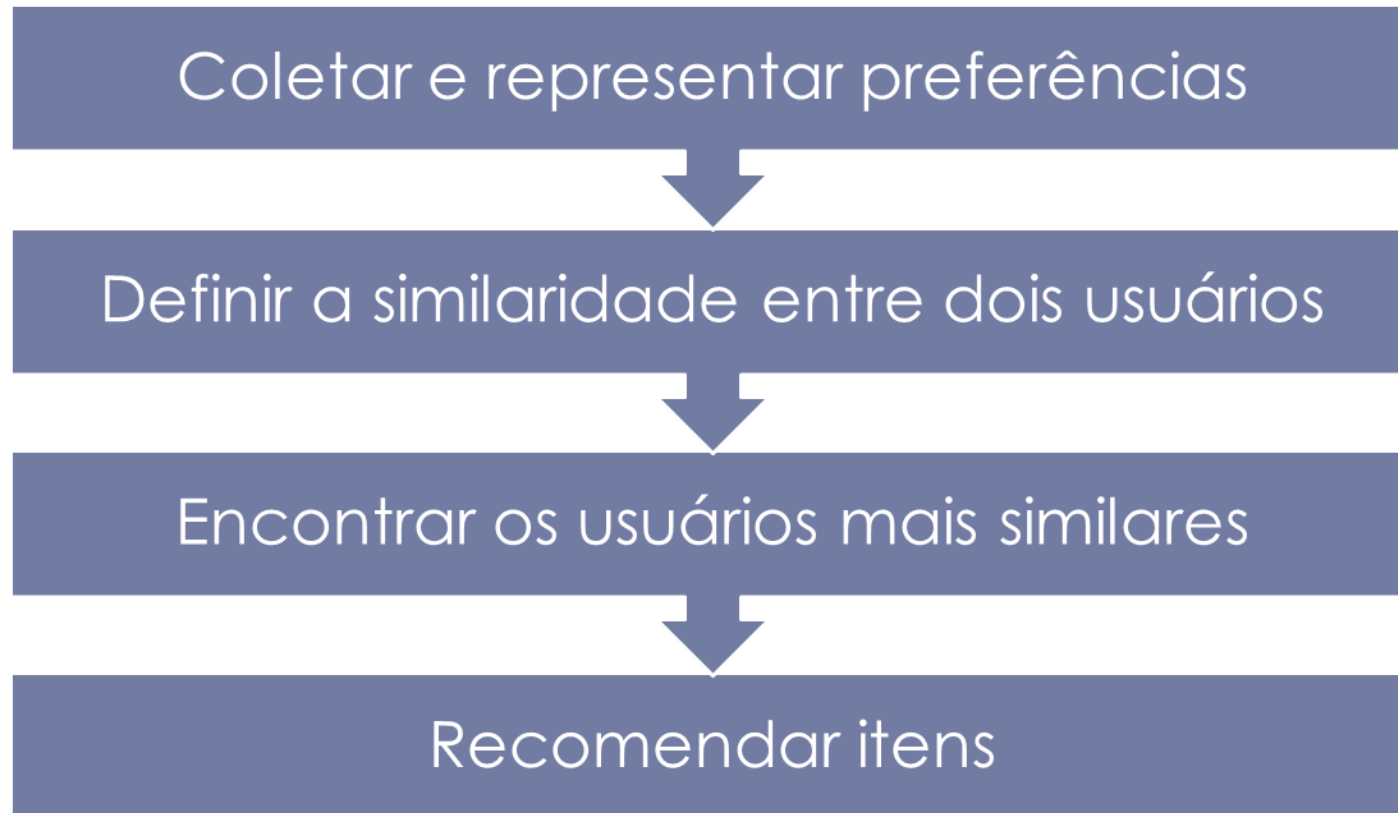
Recomendando itens

- Verificar todos os itens que um usuário avaliou
- Encontrar os itens similares
- Ponderá-los através do dicionário de itens

Movie	Rating	Night	R.xNight	Lady	R.xLady	Luck	R.xLuck
Snakes	4.5	0.182	0.818	0.222	0.999	0.105	0.474
Supeman	4.0	0.103	0.412	0.091	0.363	0.065	0.258
Dupree	1.0	0.148	0.148	0.4	0.4	0.182	0.182
Total		0.433	1.378	0.713	1.764	0.352	0.914
Normalized			3.183		2.598		2.473

```
>> reload(recommendations)
>> recommendations.getRecommendedItems(recommendations.critics, itemsim, 'Toby')
[(3.182, 'The Night Listener'),
 (2.598, 'Just My Luck'),
 (2.473, 'Lady in the Water')]
```

Algoritmo implementado



Ferramentas

- Apache Mahout

- Biblioteca para aprendizado de máquina, classificação, clustering e recomendações
- Fornece ferramentas para construir SR variados
- Trabalha sobre Apache Hadoop
 - Framework open-source para escalabilidade através de computação e armazenamento distribuídos.
- Opensource com licença Apache 2.0

- MyMediaLite

- Biblioteca leve para construção de algoritmos para sistemas de recomendação
- Opensource com licença GPL

Ferramentas

- Duine

- Framework para criar mecanismos de predição (calcula quão interessante um item é para um usuário)
- Opensource com licença LGPLv3

- GraphLab

- Originalmente escrito para tarefas de aprendizado de máquina, mas é atualmente utilizado para um grande número de tarefas de data mining
- Oferece um toolkit para filtragem colaborativa
- Opensource com licença Apache 2.0

- Lenskit

- Toolkit para construir, pesquisar e estudar sistemas de recomendação
- Opensource com licença GPL

Seção 4

Estudos de caso

Estudos de Caso

1. Netflix
2. Google AdSense

Estudo de Caso: Netflix

- Netflix: serviço de TV por internet.
- Importância da Recomendação ao Netflix:
 - Em 2006, criou uma competição que premiaria em 1 milhão de dólares quem conseguisse melhorar seu algoritmo (Netflix Prize);
 - Em 2012, 75% do conteúdo visualizado do Netflix foi proveniente de algum tipo de recomendação;
 - Maior importância: personalização das recomendações ao usuário ativo, a ordenação dos itens recomendados é o parâmetro principal.

Estudo de Caso: Netflix

- Organização do Estudo – Netflix:
 1. Netflix Prize
 2. Sistema de Recomendação Netflix
 3. Parâmetros Analisados para Recomendação
 4. Ordenação dos Itens Recomendados
 5. Conclusão

Netflix: Netflix Prize

- Netflix Prize:
 - \$1 milhão ao vencedor;
 - Critério de vitória: superar o RMSE do algoritmo existente da Netflix (0,9525) em pelo menos 10%.
 - Prêmio melhor progresso anual: \$50 mil;
 - Início em 2006 e término em 2009;
 - Vencedor do primeiro Prêmio de Progresso:
 - Equipe da Korbell
 - 8,43% de melhoria;
 - Conjunto de 107 algoritmos. Dois algoritmos aproveitados: *Matrix Factorization* e *Restricted Boltzmann Machines*.

Netflix: Netflix Prize

- Melhor Resultado: BellKor's Pragmatic Chaos
- Melhoria de 10,06%.



Leaderboard

Showing Test Score. [Click here to show quiz score](#)

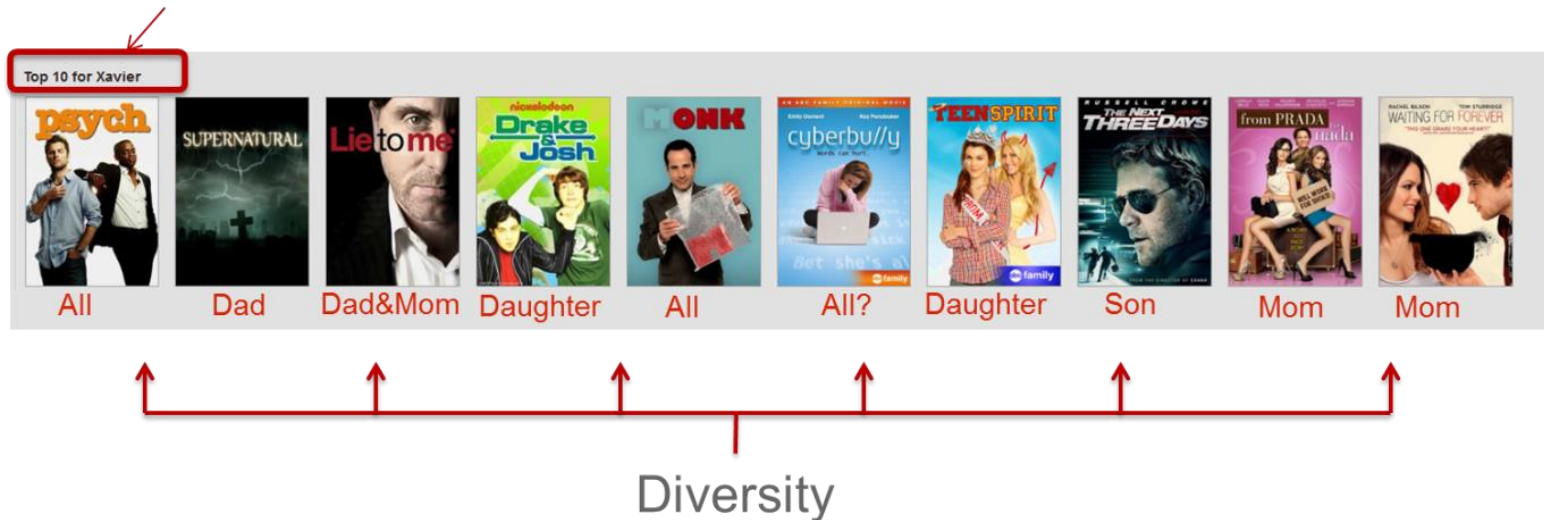
Display top leaders.

Rank	Team Name	Best Test Score	% Improvement	Best Submit Time
Grand Prize - RMSE = 0.8567 - Winning Team: BellKor's Pragmatic Chaos				
1	BellKor's Pragmatic Chaos	0.8567	10.06	2009-07-26 18:18:28
2	The Ensemble	0.8567	10.06	2009-07-26 18:38:22
3	Grand Prize Team	0.8582	9.90	2009-07-10 21:24:40
4	Opera Solutions and Vandelay United	0.8588	9.84	2009-07-10 01:12:31
5	Vandelay Industries !	0.8591	9.81	2009-07-10 00:32:20
6	PragmaticTheory	0.8594	9.77	2009-06-24 12:06:56
7	BellKor in BigChaos	0.8601	9.70	2009-05-13 08:14:09
8	Dace	0.8612	9.59	2009-07-24 17:18:43

Netflix: Sistema de Recomendação

- Sistema de Recomendação Netflix
 - Tudo se torna uma recomendação!
 - Recomendações arranjadas em grupos colocados em linhas, e cada coluna é um item do grupo.

Personalization awareness



Netflix: Parâmetros Analisados

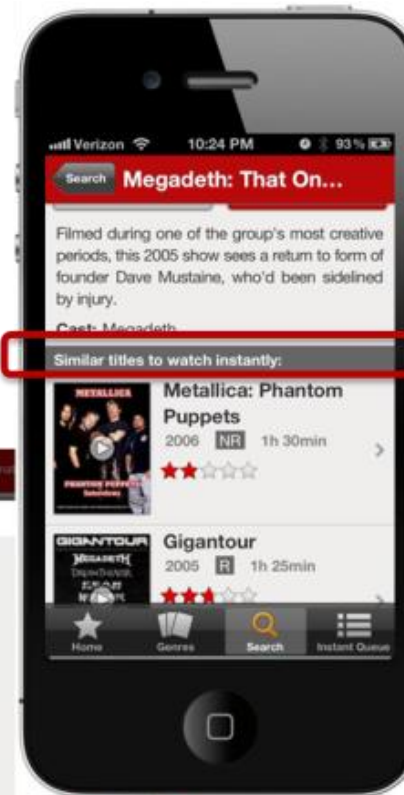
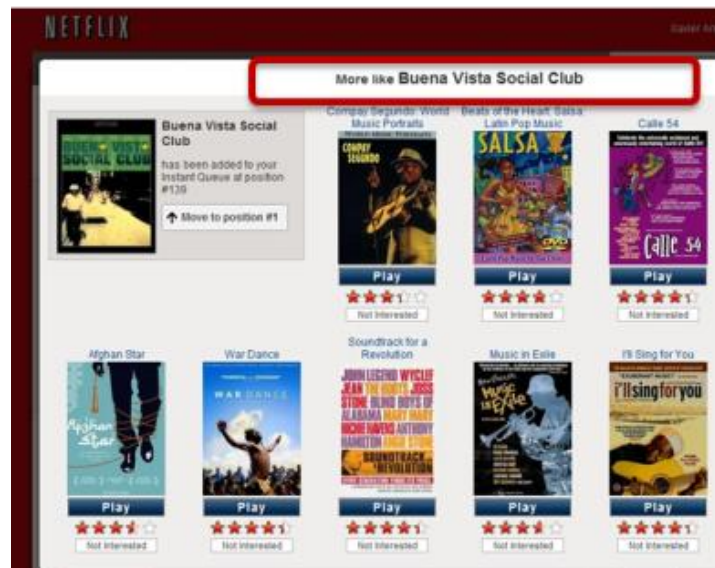
- Principais Parâmetros de Recomendação:
 - Semelhança (ou Similaridade);
 - Amigos (social);
 - Popularidade;
 - Gênero (ou Categoria);
 - Outros parâmetros podem incluir localização geográfica do usuário ou dados retirados de seu perfil ou outros acessos.
- Algoritmos observam estes parâmetros em conjunto, não separados.

Netflix: Parâmetros Analisados

Similaridade

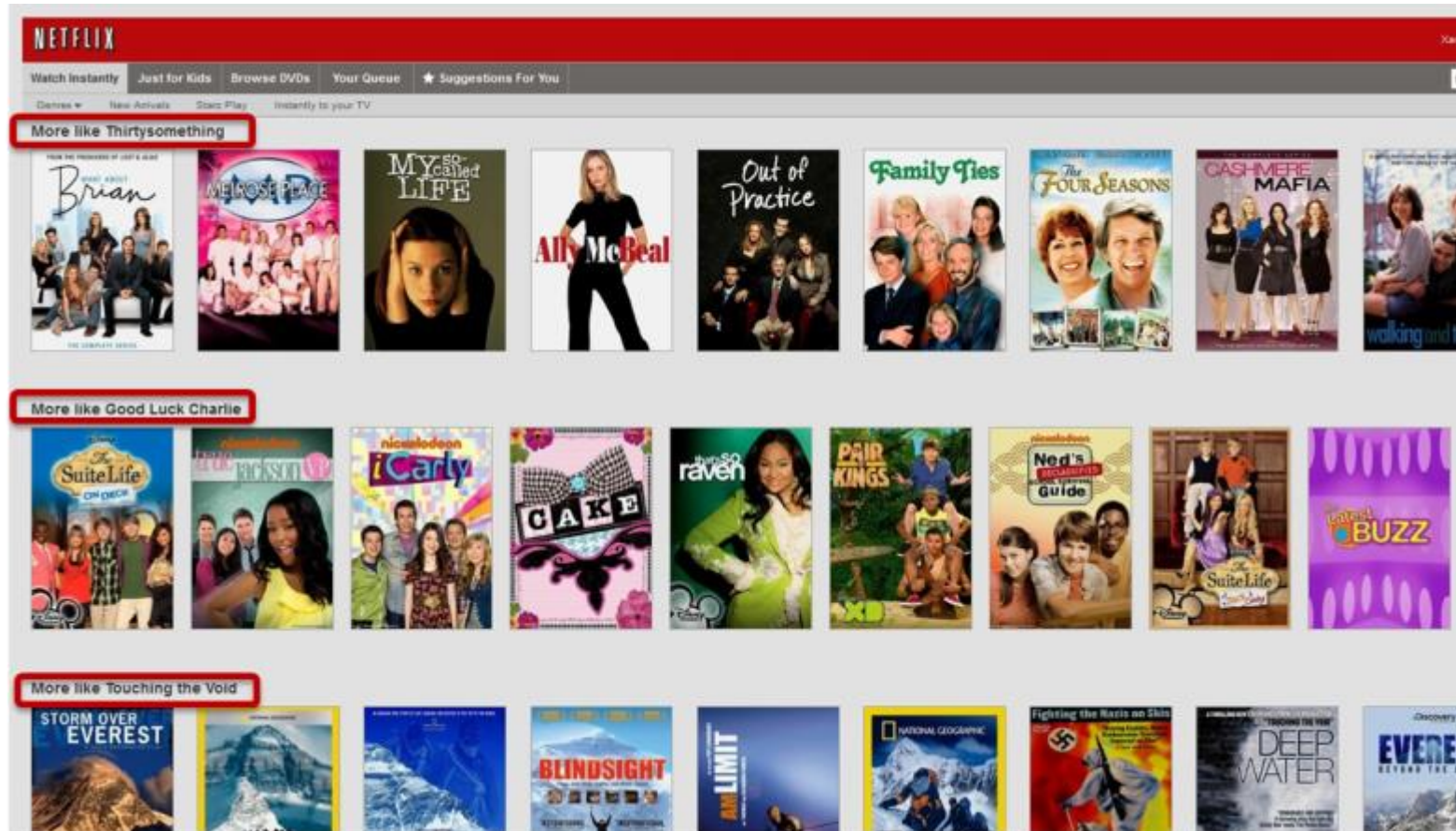
Search results

Queue Add



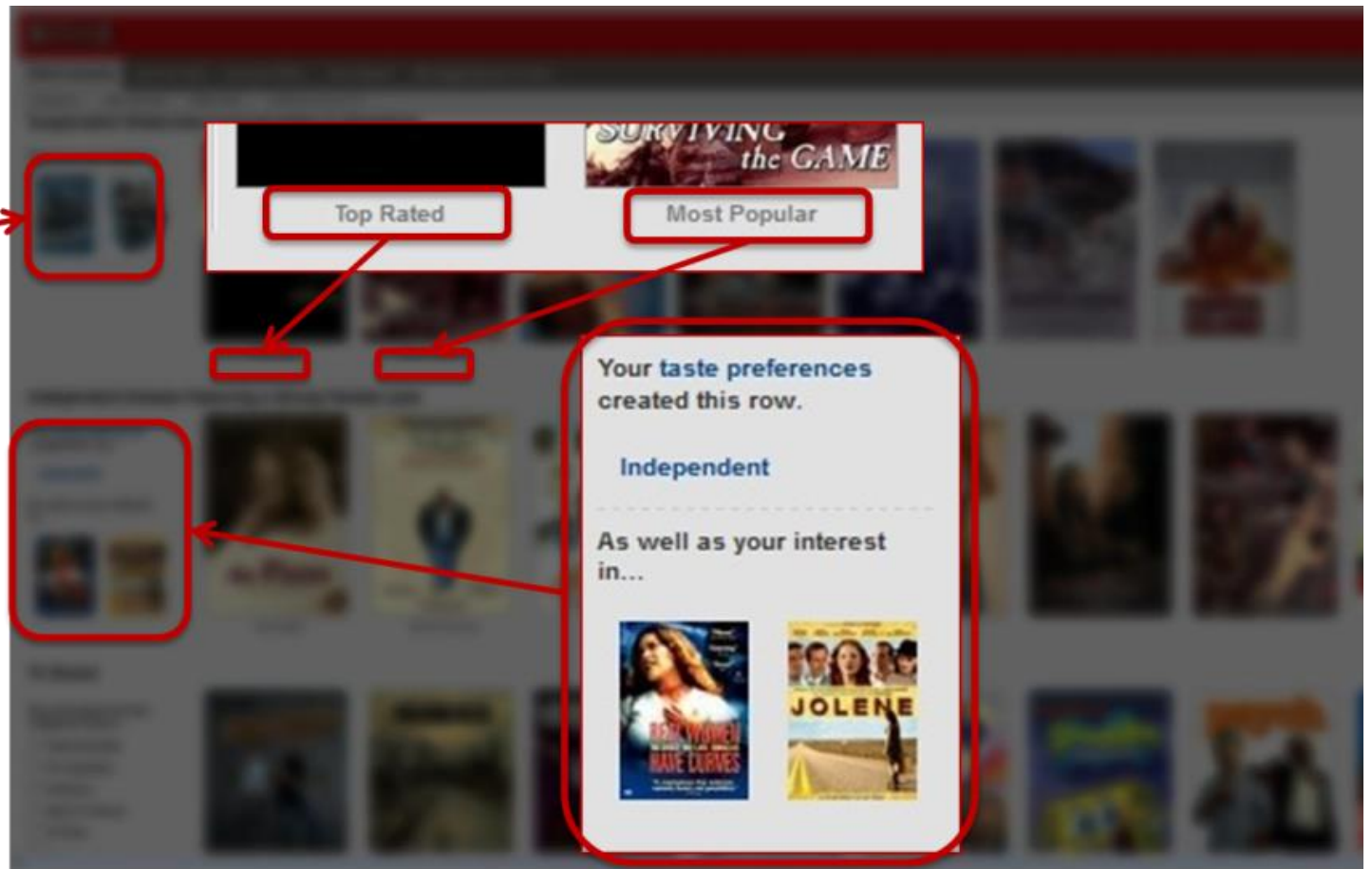
Netflix: Parâmetros Analisados

● Similaridade



Netflix: Parâmetros Analisados

- Similaridade: Gêneros Temporários

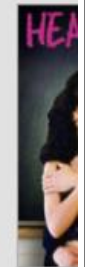
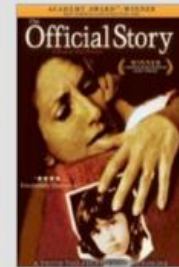


Netflix: Parâmetros Analisados

◉ Sugestões de amigos (social)

f Friends' Favorites

Based on these friends:



f Watched by your friends

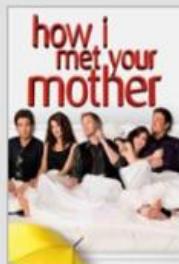
Daniel Jacobson

John Ciancutti

Mark White



mike Kail



Netflix: Parâmetros Analisados

- Popularidade:
 - Através da avaliação dos usuários para os itens assistidos;
 - Através do número de visualizações dos itens, parciais e totais;
 - Netflix possui um mecanismo para medir a popularidade de itens no Facebook, e não apenas em seu sistema.

Netflix: Parâmetros Analisado

- Gêneros:

- Trata-se da organização dos filmes em linhas determinadas por gêneros.
- Os gêneros de alto nível, mais comuns, são os que geralmente se enquadram os filmes: comédias, dramas, suspense, terror, ação, etc.
- Filmes também podem ser enquadrados em subgêneros diversos.

Netflix: Parâmetros Analisados

- Gêneros: exemplos de subgêneros

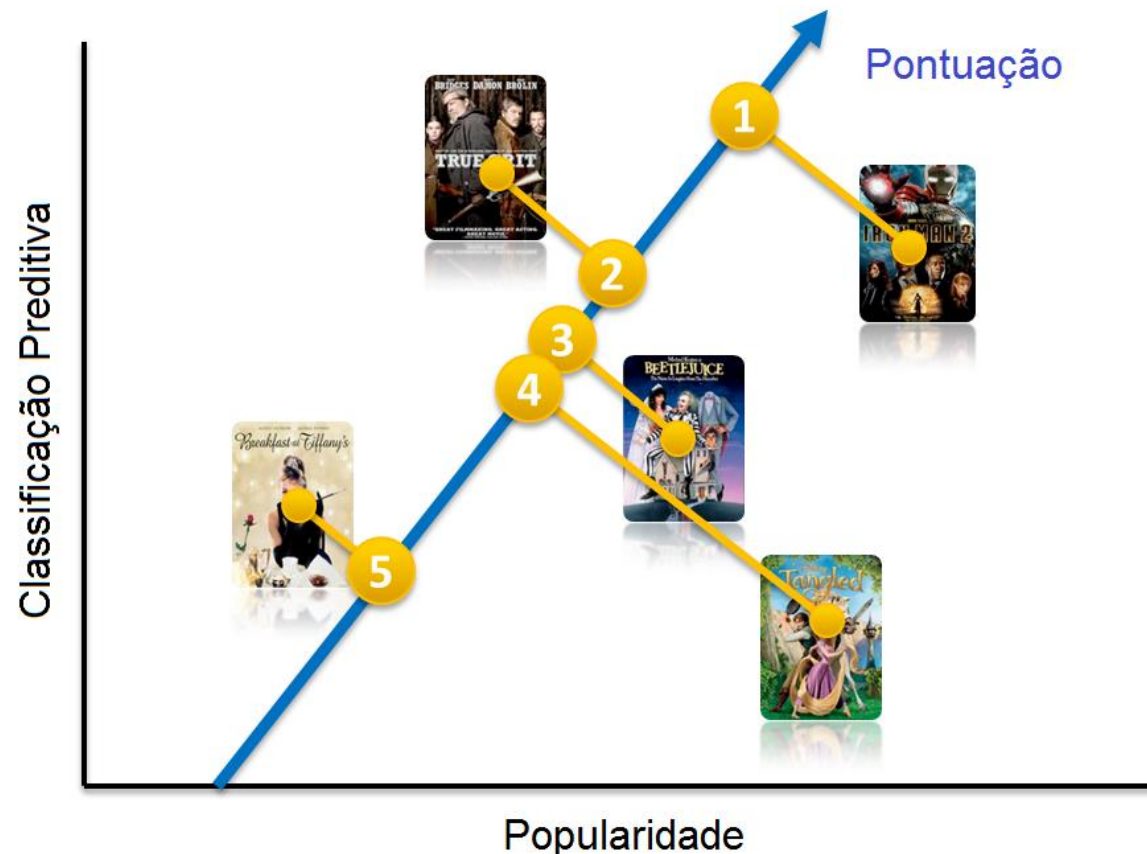


Netflix: Parâmetros Analisados

- Ordenação:
 - Encontrar a melhor ordem para exibição de itens de cada arranjo;
 - Analisa e atribui ordem de acordo com um ou mais critérios de preferência do usuário ou popularidade do item;
 - É considerada muito importante ao Netflix, pois é fator principal da personalização usuário – serviço.

Netflix: Ordenação de Itens

- Ordenação: exemplo simples de ordenação



Netflix: Conclusões

- Sistema de recomendação Netflix utilizada métodos híbridos;
- O sistema de recomendação é parte fundamental da experiência usuário – serviço, vez que boa parte da visualização de itens é por recomendação;
- A equipe Netflix busca constantemente a melhoria de seu sistema de recomendação direcionado a personalização, mostrando reconhecimento da importância deste como diferencial do serviço.

Estudo de Caso: Google AdSense

● Organização do Estudo:

1. AdSense e AdWords
2. Funcionamento do AdSense
3. Site Parceiro: Implantação e Vantagens
4. Parâmetros Analisados
5. Conclusão

AdSense e AdWords

- Google AdSense:
 - Definição;
 - Local de Aplicação;
 - Vínculo direto com o Google AdWord.
- Google AdWords:
 - Definição;
 - Funcionamento;
 - Local de Publicação.

AdSense e AdWords

Exemplo
AdWords:

Busca por
“Anúncios
do
Google”

Google Anúncios do Google

Web Imagens Mapas Shopping Vídeos Notícias Mais Ferramentas de pesquisa

Aproximadamente 105.000.000 resultados (0,40 segundos)

Anúncios relacionados a Anúncios do Google ⓘ

Google AdWords 0800 724 8845
www.google.com.br/Ads/AdWords ▾
 Comece ganhando novos clientes hoje Inscreva-se no Google AdWords!
 Google para seu Negócio tem 208.250 seguidores no Google+
 Quanto custa Segmente seus anúncios
 Histórias de sucesso Perguntas frequentes

Anuncie Sua Empresa 0 15 (48) 9991-9242
www.agenciag13.com.br/ ▾
 Encontre Seu Público Alvo, Agência Especializada em Links Patrocinados

Seu site na 1ª página já - Sua empresa no topo das pesquisas
www.awccomunicacao.com.br/adwords/ ▾
 Venda mais, Atraia novos Clientes.
 Rua Bernardo Welter, 422 - Costa e Silva, Joinville - SC - (47) 3025-8240

Anúncios do Google
www.google.com.br/ads/ ▾
 Saiba mais sobre a publicidade no Google e como ganhar dinheiro com seu site.

Google AdWords - Publicidade on-line do Google
adwords.google.com.br/ ▾
 Anuncie no Google AdWords, na seção de "Links patrocinados" ao lado dos resultados de pesquisa para incentivar o tráfego e as vendas de seu website.

Ganhe dinheiro com seu site – Editor – Anúncios do Google
www.google.com.br/ads/publisher/ ▾
 Explore o pacote completo de soluções para editores do Google criado para ...

Anúncios ⓘ

Anuncie Grátis na OLX
www.olx.com.br/desapega ▾
 Vender na OLX é Fácil.
 Fotografou, Publicou, Desapegou!

Link Patrocinado Anúncios
www.velli.com.br/ ▾
 (11) 4301-3433
 Divulge seus produtos na Internet
 Planos e Serviços para sua Empresa

Anúncios Buscadores
www.cadastrando.com.br/Anuncios-Google ▾
 0800 774 9700
 Quer Colocar Anúncios Aqui ?
 Faça Já Seu Orçamento! Confira.

Anuncie Grátis Online
www.bomnegocio.com/Anuncios ▾
 Seja Visto Por Milhares de Pessoas
 Com o Bom Negócio! Acesse e Confira

Veja seu anúncio aqui »

AdSense: Funcionamento

- Funcionamento do AdSense:
 - Baseado nos anúncios do AdWords:
 - Os dois sistemas se complementam;
 - Modelo impulsiona o crescimento do comércio eletrônico e garante lucros para a Google;
 - Entidades envolvidas:
 - Anunciante: cria o anúncio e paga sua manutenção;
 - Editor do site parceiro: hospeda anúncios em sua página da web;
 - Ferramenta Google AdSense: gerenciar os anúncios e atribui anúncios condizentes nas páginas dos parceiros.

AdSense: Site Parceiro

- Site Parceiro –Vantagens de ser parceiro:
 - Lucro de percentual por clique para o site parceiro apenas disponibilizando espaço;
 - Acesso simples e instantâneo a milhares de anunciantes;
 - Ferramenta controla os anúncios, sendo desnecessária manutenção de anúncios ou da forma que são selecionados.

AdSense: Site Parceiro

- Site Parceiro – Modo de Implantar:
 1. Cadastre-se no Google AdSense;
 2. Escolher os formatos dos anúncios;
 3. Instalar os códigos gerados pelo AdSense para sua página;
 4. Gerenciar o programa Adsense;
 5. Receber a comissão.

AdSense: Parâmetros Analisados

- Parâmetros Analisados:
 - Conteúdo da página visualizada;
 - Localização geográfica do usuário;
 - Custo optado a pagar por clique;
 - Histórico de acesso a sites (por tipo e conteúdo);
 - Informações do perfil do Google;
 - Se o usuário já interagiu com anúncios anteriormente;
 - Cookies da DoubleClick no navegador;
 - Entre outros não divulgados.

AdSense: Exemplo

Exemplo do AdSense: GameFAQs

Is giving away accounts illegal under net standards? giving away free accounts
League of Legends - PC

Is there a certain genre of games that you just flat out hate? Explain why!
Xbox 360

Do any of your cars have a chrome paintjob?
Grand Theft Auto Online - PlayStation 3

[More Message Board Topics »](#)

[game deals revealed](#)

[Metal Gear Solid 5: Ground Zeroes getting PlayStation-exclusive content](#)

[Community News Update: Tuesday 11/12](#)

[\\$10 Pokemon X/Y music collection hits iTunes](#)

[PlayStation 4 preorders ahead of Xbox One, Ubisoft suggests](#)

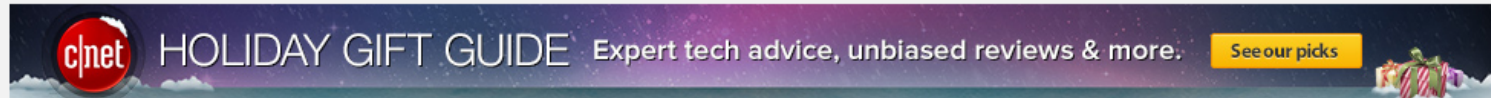
[Sony on PS4 launch -- "We're very, very confident"](#)

[Ubisoft: Watch Dogs progressing "very well," will be top-seller in 2014](#)

[Witcher dev: Publishers use DRM as "smokescreen" to cover their a**es](#)

[Rayman Legends coming to PS4, Xbox One in February](#)

[More Headlines »](#)



AdSense: Exemplo

Exemplo AdSense:
GameFAQs >
CNET

The image is a screenshot of the CNET 2013 Holiday Gift Guide page. At the top, the CNET logo and "2013 HOLIDAY GIFT GUIDE" are visible. Below this, there are several product recommendations:

- A Roku 3 streaming player with the text "OUR FAVORITES Editors' top picks for the holidays".
- A tablet displaying a game, with the text "COMPUTING ON THE GO Awesome tablets to give or receive".
- A portable Bluetooth speaker, with the text "Deal of the Day | Nov 12, 2013 Get a Braven 570 portable Bluetooth speaker for \$29.99".
- A camera, with the text "HAVE AN ADVENTURE Action cams for".

Below these recommendations, there is a Google AdSense banner for the Nexus 7 tablet, which is highlighted with a red border. The banner text reads: "Nexus 7. The world's highest resolution 7\" tablet." and includes a "Buy now" button with a price of "from \$229".

Below the banner, there is a "Categories" section with several product images and labels:

- "Tablets and e-book readers" with an image of a tablet.
- "Smartphones" with an image of a smartphone.
- "PCs and Macs" with an image of a laptop.
- "Gaming" with an image of a game console and controller.

At the bottom right, there is a "More Holiday Suggestions" section with several small product images and labels:

- "Editors' top picks" with an image of a light bulb.
- "Best tech under \$50" with an image of a remote control.
- "Tech toys for big kids" with an image of a lamp.
- "Best tech under \$100" with an image of a tablet.

AdSense: Conclusões

- Conclusões:
 - Método de recomendação baseado em conteúdo, pois nenhum dos parâmetros se baseia em algum tipo de avaliação de usuário;
 - A abordagem tem grande relevância para a Google, uma vez que quanto mais interessante e preciso se tornarem os anúncios, maior o lucro da empresa;
 - Recomendações corretas e otimizadas são essenciais para aumentar o interesse do usuário no anúncio e o número de vendas do anunciante.

Seção 5

Desafios e limitações

Desafios e Limitações

1. Esparcidade
2. The Cold-Start Problem
3. Fraude
4. Filter Bubble

Esparcidade

- A maioria dos usuários não avalia grande parte dos seus itens;
- Isso diminui a probabilidade de encontrar usuários com avaliações parecidas (filtragem colaborativa).

$$\begin{bmatrix} 1.1 & 0 & 0 & 0 & 0 & 0 & 0.5 \\ 0 & 1.9 & 0 & 0 & 0 & 0 & 0.5 \\ 0 & 0 & 2.6 & 0 & 0 & 0 & 0.5 \\ 0 & 0 & 7.8 & 0.6 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1.5 & 2.7 & 0 & 0 \\ 1.6 & 0 & 0 & 0 & 0.4 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0.9 & 1.7 \end{bmatrix}$$

The Cold-Start Problem

- Este problema pode ocorrer quando há itens e usuários novos;
- Surge principalmente em sistemas de filtragem colaborativa, pois os itens precisam de avaliação anteriores;
- Este problema também se aplica a itens “obscuros”, ou seja, um item dos menos requisitados

Fraude

- Em sistemas que dependem de avaliações de produtos, vendedores podem manipular o sistema;
- Essa manipulação ocorre utilizando perfis falsos que avaliam os competidores de forma negativa e o vendedor fraudulento de forma positiva.

Filter Bubble

- Fenômeno em que um usuário acaba sendo isolado de outras idéias, culturas e produtos potencialmente benéficos;
- Ocorre devido à personalização criada pelo sistema de recomendação, que recomenda itens novos baseados nos anteriores, ou em certos perfis.

Considerações Finais

- Sistemas de recomendação podem ser uma ótima forma de aumentar o lucro ou relevância de um serviço online;
- Possuem vários benefícios para os usuários e prestadores de serviço, mas é necessário tomar cuidado para não infringir a privacidade do usuário nem omitir itens disponíveis no sistema.

Referências

AJUDA do AdSense. Disponível em:

<<https://support.google.com/adsense/troubleshooter/1631343>>
. Acesso em: 11 nov. 2013.

AMATRIAIN, X; BASILICO, J. Netflix Recommendations: Beyond the 5 stars. 2012. Disponível em:

<<http://techblog.netflix.com/2012/04/netflix-recommendations-beyond-5-stars.html>>. Acesso em: 11 nov. 2013.

APACHE Mahout. Disponível em: <<http://mahout.apache.org/>>.

Acesso em: 10 de novembro de 2013.

DUINE. Disponível em:

<<http://www.duineframework.org/index.html>>. Acesso em: 10 de novembro de 2013.

Referências

GROSSMAN, L. How Computers Know What We Want — Before We Do. Disponível em:
<<http://www.time.com/time/magazine/article/0,9171,1992403,00.html>>. Acesso em: 20 ago. 2013.

GATTO, E. C. Palestra unesp 2012, 15 out. 2012. Disponível em:
<<http://www.slideshare.net/elainececiliagatto/palestra-unesp-2012>>. Acesso em: 10 out. 2013

FELIPINI, D. Adwords e Adsense, os filhos gêmeos do Google. Disponível em: < <http://www.e-commerce.org.br/artigos/adsense-adwords.php> >. Acesso em: 11 nov. 2013.

FELIPINI, D. Google AdSense. Disponível em:
<http://www.e-commerce.org.br/google_adsense.php>. Acesso em: 11 nov. 2013.

Referências

GraphLab. Disponível em: <<http://graphlab.org/>>. Acesso em: 10 de novembro de 2013.

INGERSOLL, G. Introdução ao Apache Mahout (2009). Disponível em: <<http://www.ibm.com/developerworks/br/java/library/j-mahout/>>. Acesso em: 15 de outubro de 2013.

LENSKIT. Disponível em: <<http://lenskit.grouplens.org/>>. Acesso em: 10 de novembro de 2013.

MELVILLE, P.; SINDHWANI, V. Encyclopedia of machine learning. [s.l.] Springer-Verlag, chapter Recommender systems, 2010.

MILLER, E.; How Not To Sort By Average Rating. Disponível em: <<http://www.evanmiller.org/how-not-to-sort-by-average-rating.html>>. Acesso em: 9 out. 2013.

Referências

MYMEDIALITE. Disponível em: <<http://mymedialite.net/>>. Acesso em: 10 de novembro de 2013.

NETFLIX. Disponível em: <<http://pt.wikipedia.org/wiki/Netflix>>. Acesso em: 11 nov. 2013.

O QUE é o Google AdWords?. Disponível em: <<http://www.clinks.com.br/videos-tutoriais/o-que-e-google-adwords/>>. Acesso em: 11 nov. 2013.

PARISER, E. The filter bubble: What the Internet is hiding from you. Penguin (UK). 2011.

Referências

RICCI, F.; ROKACH, L.; SHAPIRA, B. Introduction to Recommender Systems Handbook. In: RICCI, F. et al. (Eds.). Recommender Systems Handbook. Boston, MA: Springer US, 2011. p. 1–35.

SAMPAIO, I.; Aprendizagem Ativa em Sistemas de Filtragem Colaborativa. 2007

SCHRODER, G.; THIELE, M.; LEHNER, W. "Setting Goals and Choosing Metrics for Recommender System Evaluations." UCERSTI Workshop at the ACM Conference on Recommender Systems, 2011.

Segaran, Toby. Programming Collective Intelligence: Building Smart Web 2.0 Applications. Beijing: O'Reilly, 2007.