

Sistemas Colaborativos

Aula: Sistema de recomendação

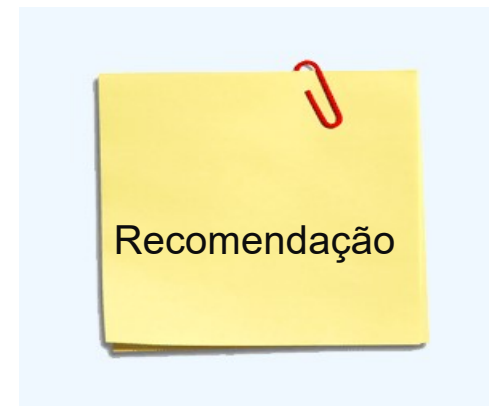
Introdução

- Muitas msgs!
- Muitos artigos!
- Muitos filmes!
- Muito conteúdo!
- Recomendação



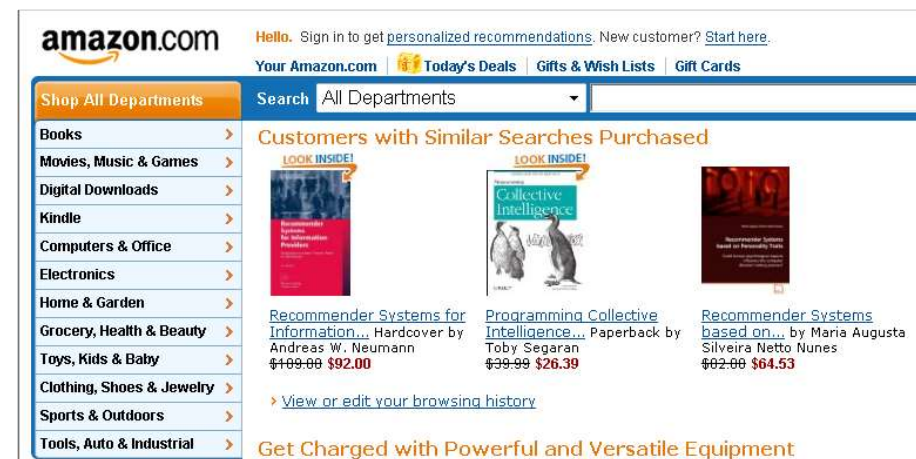
Introdução

- Os sistemas de recomendação têm por objetivo:
 - Reduzir a sobrecarga de informação
 - Realizando filtragens de itens
 - Baseadas no perfil de interesses do usuário



Introdução

- Definição
 - Sistemas de recomendação (SR) são filtros de informação para apresentar itens (filmes, músicas, livros, webpages..) que muito provavelmente serão do interesse do usuário. [Shaffer 2001].
 - SRs têm sido utilizados, em site de compra, por exemplo: Amazon e Submarino, na intenção de sugerir produtos que agradem ao cliente.

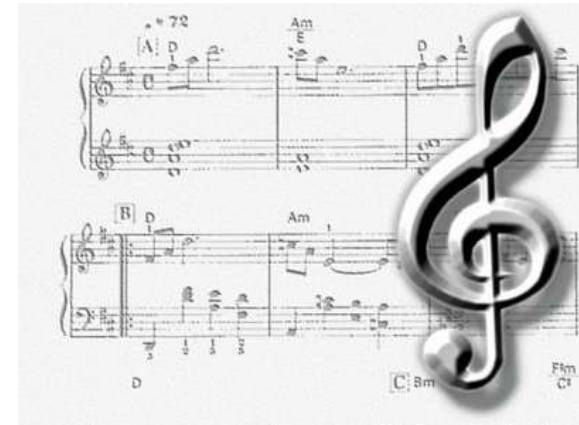


Introdução

- Principais componentes de um SR
 - Usuários
 - Conteúdo (itens)
- Recomendação → função da relação entre usuário e itens
- Formalização
 - Seja
 - Um usuário u ,
 - Um conjunto C de conteúdo (itens), em que $c_i \in C$,
 - A recomendação “ c ” é feita tal que:
 - $F(u, c) = \text{MAX } F(u, c_i)$,
 - F é uma função que determina a relevância de c em relação a u

Introdução

- Exemplos clássicos:
 - Ringo [MIT; Shardanand e Maes 1995]
 - *Recomendação de músicas*
 - Avaliação inicial de material (músicas) gera perfil inicial do usuário;
 - Recomendações geradas para o usuário;
 - Usuário avalia a sugestão, ajuste de perfil



- Fab System [Balabanovic 1997]

- Recomendação de documentos
- Abordagem híbrida

- Filtragem colaborativa: identifica as correlações e faz recomendações
- Filtragem baseada em conteúdo: combina o conhecimento e a experiência pessoal para criar o perfil dos usuários e identificar as similaridades.

C2 cotidiano TERÇA-FEIRA, 23 DE JANEIRO DE 2007 FOLHA DE SÃO PAULO

RUBEM ALVES

O aluno perfeito

ERA UMA vez um jovem casal que estava muito feliz. Ela estava grávida, e eles esperavam com grande ansiedade o filho que iria nascer.

Transcorridos os nove meses de gravidez, ele nasceu. Ela deu à luz um lindo computador! Que felicidade ter um computador como filho! Era o filho que desejavam ter! Por isso eles haviam resado muito durante toda a gravidez, chegando mesmo a fazer promessas.

O batizado foi uma festa. Deram-lhe o nome de Memorioso, porque julgavam que uma memória perfeita é o essencial para uma boa educação. Educação é memorização. Crianças com memória perfeita vão bem na escola e não têm problemas para passar no vestibular.

E foi isso mesmo que aconteceu. Memorioso memorizava tudo que os professores ensinavam. Mas, tudo mesmo. E não reclamava. Seus com-

panheiros reclamavam, diziam que aquelas coisas que lhes eram ensinadas não faziam sentido. Suas inteligências recusavam-se a aprender. Tiravam notas ruins. Ficavam de recuperação.

Isso não acontecia com Memorioso. Ele memorizava com a mesma facilidade a maneira de extrair raiz quadrada, reações químicas, fórmulas de física, acidentes geográficos, populações de países longínquos, datas de eventos históricos, nomes de reis, imperadores, revolucionários, santos, escritores, descobridores, cientistas, palavras novas, regras de gramática, livros inteiros, línguas estrangeiras. Sabia de cor todas as informações sobre o mundo cultural.

Ele se chamava Memorioso, pois seus pais julgavam que a memória perfeita é essencial para uma boa educação.

A memória de Memorioso era igual à do personagem do Jorge Luis Borges de nome Funes. Só tirava dez, o que era motivo de grande orgulho para os seus pais. E os outros casais, pais e mães dos colegas de Memorioso, morriam de inveja. Quando filhos chegavam em casa trazendo boletins com notas em vermelho eles gritavam: "por que você não é como o Memorioso?"

Memorioso foi o primeiro no vestibular. O cursinho que ele frequentava-

ram publicou sua fotografia em outdoors. Apareceu na televisão como exemplo a ser seguido por todos os jovens. Na universidade, foi a mesma coisa. Só tirava dez. Chegou, finalmente, o dia tão esperado: a formatura. Memorioso foi o grande herói, elogiado pelos professores. Ganhou medalhas e mesmo uma bolsa para doutoramento no MIT.

Depois da cerimônia acadêmica foi a festa. E estavam todos felizes no jantar quando uma moça se aproximou de Memorioso e se apresentou: "Sou repórter. Posso lhe fazer uma pergunta?" "Pode fazer", disse Memorioso confiante. Sua memória continha todas as respostas.

Al ela falou: "De tudo o que você memorizou: qual foi aquilo que você mais amou? Que mais prazer lhe deu?"

Memorioso ficou mudo. Os circuitos de sua memória funcionavam com a velocidade da luz procurando a resposta. Mas aquilo não lhe fora ensinado. Seu rosto ficou vermelho. Começou a suar. Sua temperatura subiu. E, de repente, seus olhos ficaram muito abertos, purados, e se ouviu um chiado estranho dentro de sua cabeça, enquanto fumava sua por suas orelhas. Memorioso primeiro travou. Deixou de responder a estímulos. Depois apagou, entrou em coma. Levado às pressas para o hospital de computadores, verificaram que seu disco rígido estava irreparavelmente danificado.

Há perguntas para as quais a memória não tem respostas. E que tais respostas não se encontram na memória. Encontram-se no coração, onde mora a emoção.

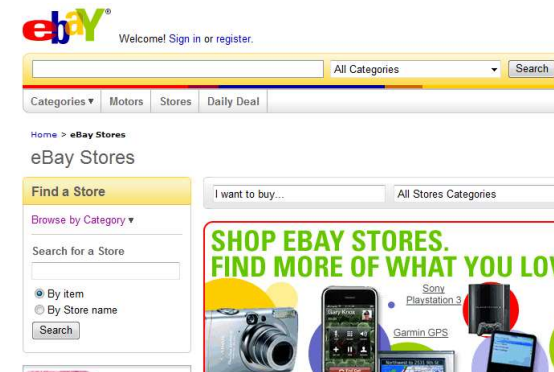
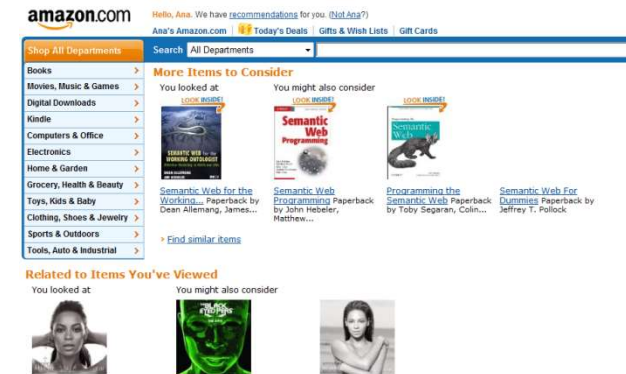
Introdução

- Exemplos clássicos:
 - GroupLens (universidade de Minnesota; Konstan 1997)
 - Recomendação de produtos
 - Computadas as aval. Dos usuários (valores de 01 até 05) para uma futura comparação entre as avaliações
 - Busca da melhor correlação entre avaliações
- Drugstore.com
 - Farmácia on-line onde os clientes indicam suas preferencias quando compram um produto de determinada categoria.
 - Preenche-se um formulário (idade e os sintomas apresentados)
 - Por meio dessas informações o sistema retorna uma lista de produtos recomendados que atendam às necessidades do cliente



Introdução

- Amazon.com
 - Avaliação de produtos variados
 - Vários tipos de técnicas e recomendações dadas
- E-Bay
 - Avaliação de transações
 - Vários tipos de técnicas e recomendações dadas
 - “Direito de Resposta”



Introdução

- Taxonomia de SR
- 3 categorias de classificação
 - Entrada e Saída funcional
 - Entrada sobre o consumidor alvo
 - Entrada da comunidade
 - Saída
 - Itens de Projeto
 - Grau de personalização
 - Entrega ou liberação
 - Método de recomendação



Parte 1

TIPOS DE SISTEMAS DE RECOMENDAÇÃO BASEADOS NO CONTEÚDO DE ENTRADA E SAÍDA

Categoria entrada e saída funcional

- Conteúdo de entrada e saída
 - Entrada oriunda do consumidor alvo
 - Entrada da comunidade
 - Saída

Entrada oriunda do consumidor alvo

- Navegação implícita na qual o SR analisa que páginas o consumidor visitou até agora
 - Exemplo: “Na página que você acabou de visitar ...”
- Navegação explícita caracterizada pelo fato do consumidor ter visitado uma determinada página
- Consultas prévias do usuário, das quais o SR pode extrair palavras-chave utilizadas ou atributos consultados
- Avaliação de produto ou serviço pelo usuário
- Histórico de compras



Entrada da comunidade

- Atributos de itens
- Histórico de compras da comunidade
- Avaliações explícitas
- Comentários dos usuários
- Popularidade externa dos itens, do tipo “mais vendidos”, “dez mais”, “cem melhores”



Saídas

- As saídas dos sistemas de recomendação são:
 - Sugestões
 - Lista (estatística) → Top-N
 - Avaliações
 - Resenhas

Frequently Bought Together

Customers buy this book with [Data Mining: Practical Machine Learning Management Systems](#) by Ian H. Witten



Price For Both: **\$69.59**

 Add both to Cart

 Add both to Wish List

[Show availability and shipping details](#)

Most Helpful Customer Reviews

6 of 7 people found the following review helpful:

★★★★★ **What you really need to know about search engines..**

By [T. Bell](#) (Christchurch, New Zealand) - [See all my reviews](#)

Anyone who has used a search engine - and who hasn't? - should read this book. It provides a coherent look at how search engines work, and their role in our information society. It is sobering: for many people, access to the internet is through one site "dragon" guarding access to a mine of information.

I particularly like the writing style; the (somewhat dry) humour and the tools that we use daily are shown in a new light. The book is suitable for the technically inclined. It provides details about how search engines work and illustrations, as well as the exposing the social implications.

Some of the important issues covered include the borderline between determining the authority of web pages compared with their popularity and access to information. Topics range from the great library of Alexandria to "Britney Spears" typed into Google.

This book looks set to become part of the computing canon, and would be a good book on a coffee table. You won't be able to use it to implement your ideas at that level of detail. But it's a thought-provoking read, and would be a good book for anyone inclined people in your life.

 [Comment](#) | [Permalink](#) | Was this review helpful to you?



Web Dragons: Inside the Myths of Search Engine Technology Information and Systems by Ian H. Witten, Marco Gori, and Te

Buy new: ~~\$35.95~~ **\$28.04** [30 Used & new](#) from **\$9.97**

Get it by **Thursday, Oct 1** if you order in the next **4 hours** and choose one-Eligible for **FREE** Super Saver Shipping.

★★★★★ (4)

Excerpt - page 219: "... provide the underlying mechanics. A type of program (books, news, music) that may interest a user ..."

Surprise me! [See a random page](#) in this book.

Books: [See all 835 items](#)

Parte 2

TIPOS DE SISTEMAS DE RECOMENDAÇÃO BASEADOS CATEGORIA DE ITENS DE PROJETO

Categoria itens de projeto

- Entrega ou liberação
- Grau de personalização

Entrega ou liberação

- Entrega pró-ativa (empurrão ou “push”)
 - O sistema busca consumidores de recomendações, quase “spam”

Special Offers and Product Promotions

- Subscribe to our [weekly newsletter](#) to learn about free downloads, special deals, and new releases.
-

- Entrega sob demanda (“pull”)
 - Os usuários controlam quando desejam ver recomendações



- Entrega passiva
 - Recomendação surge no contexto de uma navegação

Customers Who Bought This Item Also Bought

Page 1 of 17



Grau de personalização

- Grau de personalização
 - Não personalizados
 - Recomendações genéricas
 - Efêmeros
 - Comportamentos correntes ou recentes
 - Persistentes
 - Utilizam-se de todo o Banco de Dados de aquisições e comportamentos de usuários



Customers Who Bought This Item Also Bought



Ana's Amazon.com > Recommended for You
(If you're not Ana Cristina Bicharra Garcia, click here.)

Just For Today

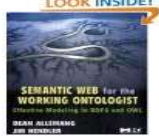
[Browse Recommended](#)

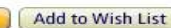
Recommendations

[All Electronics](#)
[Baby](#)
[Beauty](#)
[Books](#)
[Camera & Photo](#)
[Clothing & Accessories](#)
[Computer & Accessories](#)
[Gourmet Food](#)
[Grocery](#)

These recommendations are based on [items you own](#) and more.

view: [All](#) | [New Releases](#) | [Coming Soon](#)

1.  **LOOK INSIDE!** **Semantic Web for the Working Ontologist: Effective Modeling in RDFS and OWL**
by Dean Allemang (May 9, 2008)
Average Customer Review:  (11)
In Stock
List Price: \$49.95
Price: \$44.95
33 used & new from \$36.70
☐ I own it ☐ Not interested  Rate this item
Recommended because you purchased **Model Driven Architecture and Ontology Development** and more ([Fix this](#))

Parte 3

TIPOS DE SISTEMAS DE RECOMENDAÇÃO BASEADOS NOS MÉTODOS DE GERAÇÃO DA RECOMENDAÇÃO

Categoria método de recomendação

- Recomendação baseada em busca
 - Recuperação direta é feita por consultas tradicionais
 - Seleção manual é feita pela navegação através de estruturas hierárquicas baseadas em categorias
 - Sumários estatísticos compreendem as medidas de popularidade dentro de uma comunidade, os sumários de avaliações e as agregações de avaliações
- Recomendação baseada em filtragem
 - Filtragem baseada em conteúdo
 - Filtragem Colaborativa
- Descoberta de conhecimento
 - Regras de associação
 - Classificadores
 - Agrupamentos

Recuperação direta

- Usuário especifica consulta→Sistema recupera itens



Recuperação direta

- Vantagens
 - Simplicidade de implementação
- Desvantagens
 - Não ser muito poderoso
 - Depende do critério de classificação dos itens recuperados
 - Não recomenda nada além do que foi pedido

Recuperação baseada em categorias

- Informações classificadas em categorias hierárquicas
- Ao encontrar o nível de detalhamento de categoria desejado, o usuário aciona a solicitação da lista de itens disponíveis



Any Category Algorithms Anthropology **Artificial Intelligence** Business & Culture Databases Education
Engineering Finance Graphic Design Graphics & Multimedia History Humanities Industries & Professions
Internet Management & Leadership Networks, Protocols & APIs O'Reilly Operating Systems Philosophy Physics
Political Science Politics Psychology **Psychology & Counseling** Pure Mathematics

Recuperação baseada em categorias

- Vantagens
 - Simplicidade de implementação.
- Desvantagens
 - Não recomenda nada além do que foi pedido
 - Altamente dependente da hierarquia de categorias
 - Se for muito específico será ineficiente
 - Se for pouco específico as recomendações não serão relevantes

Recomendação baseada em estatística

- Mais vendidos
- Mais procurados
- Melhor avaliado

The image shows a screenshot of the Amazon.com website. At the top right, there is a 'TOP 5 LIVROS' (Top 5 Books) section. A red circle highlights this section, and a red arrow points from it to the text 'Sumário Estatístico' (Statistical Summary). Below this, another red arrow points from the same text to a red circle around the first recommendation in the 'Recommended for You' section. The first recommendation is the book 'Semantic Web for the Working Ontologist: Effective Modeling in RDFS and OWL' by Dean Allemang, published May 9, 2008. It has an average customer review of 4.5 stars (11 reviews) and is in stock. The list price is \$49.95, and the current price is \$44.95. The 'Recommended for You' section also includes links for 'Just For Today', 'Browse Recommended', and 'All Electronics', 'Apparel & Accessories', 'Baby', and 'Books'.

Sumário Estatístico

Sumário Estatístico

Shop All Departments Search All Departments Ana's Amazon.com Your Browsing History Recommended For You Rate These Items Improve Your Recommendations Your Profile Your Communities

Ana's Amazon.com > Recommended for You
(If you're not Ana Cristina Richarra Garcia, click here.)

Just For Today
[Browse](#)
[Recommended](#)

These recommendations are based on [items you own](#) and more.
view: [All](#) | [New Releases](#) | [Coming Soon](#)

Recommendations
[All Electronics](#)
[Apparel & Accessories](#)
[Baby](#)
[Books](#)

1. **LOOK INSIDE!** **Semantic Web for the Working Ontologist: Effective Modeling in RDFS and OWL** by Dean Allemang (May 9, 2008)
Average Customer Review: (11)
In Stock
List Price: \$49.95
Price: \$44.95
[Add to Cart](#) [Add to Wish List](#)

Recuperação baseada em categorias

- Vantagens
 - Simplicidade de implementação.
- Desvantagens
 - Recomendações genéricas

FILTRAGEM COLABORATIVA

Filtragem colaborativa

- Filtragem colaborativa (FC)
 - Passo 1: Calcular a similaridade entre os usuários (perfil **similar**). Em FC usa-se a avaliação dada pelos usuários aos produtos como o parâmetro para calcular a similaridade entre os usuários. Distância= 1-correlação
 - Técnicas para calcular correlação entre 2 usuários
 - Pearson
 - Distância Euclidiana
 - Regra do Coseno
 - Regra do Coseno ajustado
 - Passo 2: Selecionar os vizinhos mais próximos
 - Selecionar os N vizinhos mais próximos ou distantes de até x
 - Passo 3: Fazer previsão sobre avaliação do usuário alvo para um produto não consumido.

Filtragem colaborativa

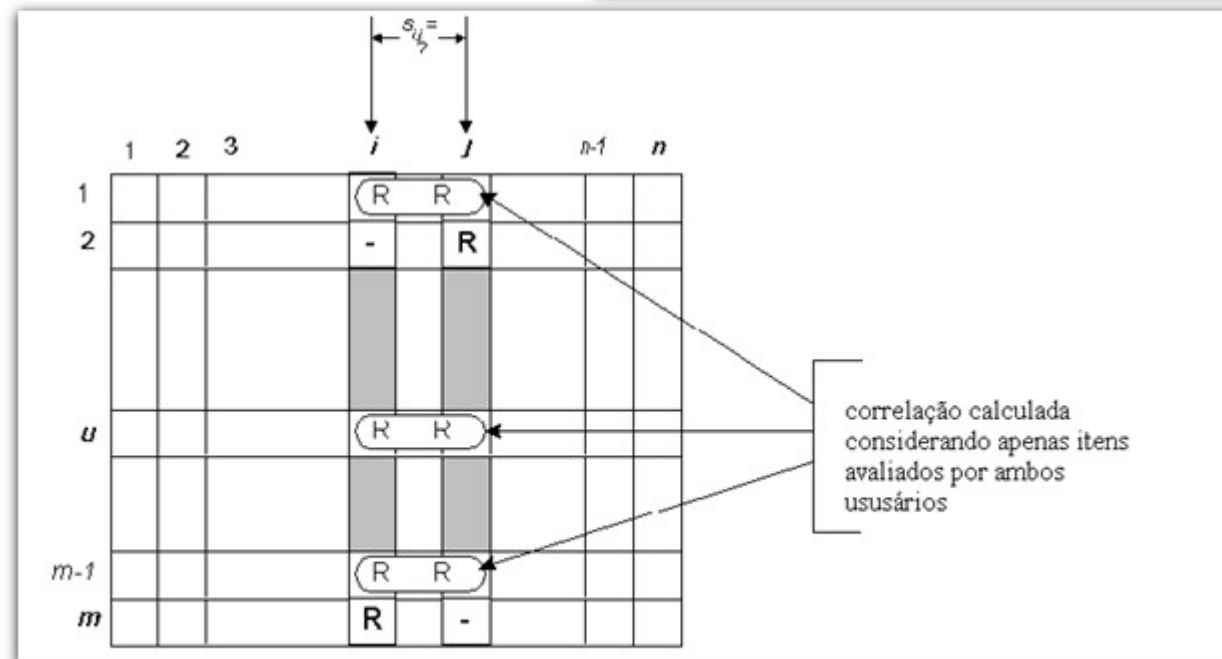
- Passo 1:
 - Correlação entre usuários
 - Pearson
 - Euclidiana
 - Co-seno

Correlação de Pearson

$$sim(i, j) = corr_{i,j} = \frac{\sum_{u \in U} (R_{u,i} - \bar{R}_i)(R_{u,j} - \bar{R}_j)}{\sqrt{\sum_{u \in U} (R_{u,i} - \bar{R}_i)^2} \sqrt{\sum_{u \in U} (R_{u,j} - \bar{R}_j)^2}}$$

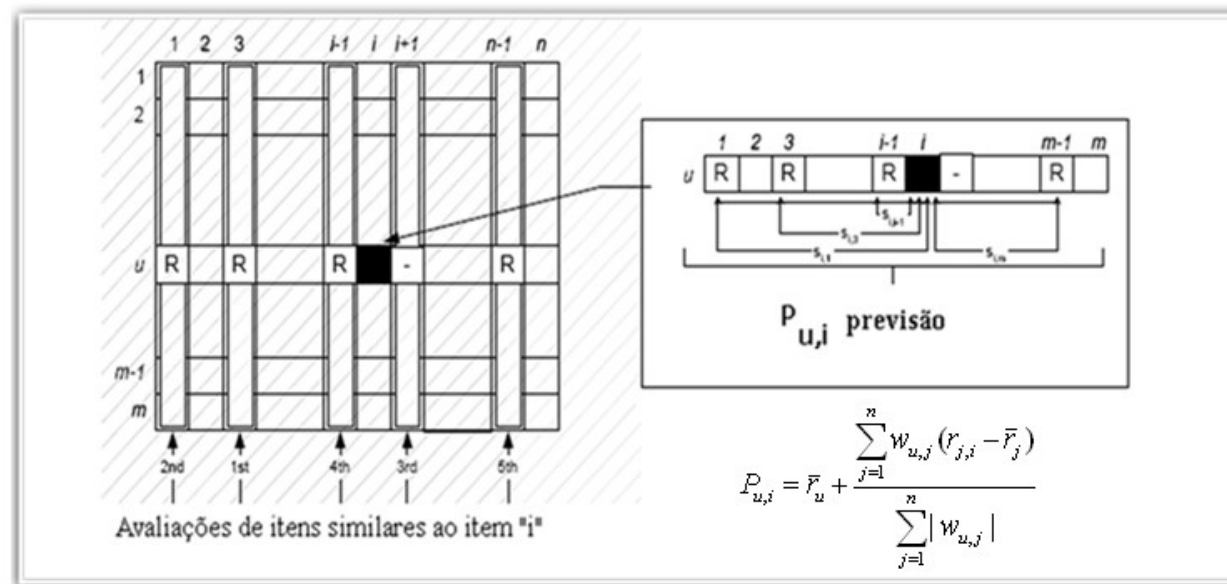
$R_{u,i}$ é a avaliação do usuário "u" no item "i"

$\bar{R}_i$ é a média de avaliações do i-th item



Filtragem colaborativa

- Passo 2: Ordenar & Selecionar os N vizinhos mais próximos do alvo
- Passo 3: Prever a avaliação do usuário alvo ao item selecionado



Filtragem colaborativa

- Exemplo: considere a base de filmes avaliados da locadora XPTO

	Maria	José	João	Pedro	Laura	Manoel	Jaime	Rafael	Lucas	Reinaldo	Felipe	Eva
Gladiator (2000)	4	5	2	3	3	4	4	3	3	4	3	4
Braveheart (1995)	4	4	2	4	3	4	5	2	3	4	3	4
Hannibal (2001)	3	5	2	3	3	5	5	2	3	3	3	3
Black Hawk Down (2001)	4	4	3	3	3	5	4	2	3	4	3	3
The Matrix (1999)	4	4	3	4	3	4	4	2	3	4	3	4
Pulp Fiction (1994)	3	4	3	3	4	5	4	3	3	4	4	4
Training Day (2001)	4	4	3	3	4	4	4	3	3	4	4	3
GangsOf New York (2002)	4	4	3	4	4	5	5	2	3	4	4	3
Pearl Harbor (2001)	4	5	3	4	4	4	5	2	3	3	4	3
Spawn (1997)	3	4	2	3	3	4	4	3	2	3	3	3
TheOriginalKings Comedy	2	3	4	2	4	2	2	5	5	3	3	4
Ice Age (2002)	3	2	5	3	3	3	3	5	5	2	4	3
Finding Nemo (2003)	3	3	4	2	3	2	3	4	5	2	4	4
Zoolander (2001)	3	3	4	2	4	3	2	5	4	3	3	4
Kangaroo Jack (2003)	2	2	4	3	4	3	2	4	5	3	4	4
Bruce Almighty (2003)	2	2	4	3	3	3	2	4	4	3	4	4
Scary Movie (2000)	2	3	4	2	4	3	3	4	5	2	4	3
Bad Boys II (2003)	2	2	5	3	4	3	2	4	4	2	4	4
Men in Black (1997)	3	3	5	2	4	2	2	5	4	3	4	4
Nixon (1995)	5	4	3	3	2	3	4	3	2	3	1	1
The Pianist (2002)	4	3	2	4	2	3	4	2	3	4	2	1
Hamlet (1996)	5	4	3	3	2	3	3	2	3	4	2	1
Seven (1995)	4	3	2	3	1	3	4	2	2	4	1	1
Malena (2000)	5	4	3	3	1	4	3	3	3	3	2	2

Filtragem colaborativa

- Continuação

	Maria	José	João	Pedro	Laura	Manoel	Jaime	Rafael	Lucas	Reinaldo	Felipe	Eva
Read Review	5	3	2	4	1	4	4	3	3	3	2	1
SurvivingPicasso	5	4	2	3	2	4	3	2	2	4	2	2
A Few Good Men (1992)	4	3	3	3	1	3	4	2	3	4	2	2
TheMothman Prophecies	5	4	3	4	2	4	4	3	2	3	1	2
Goodbye Lenin! (2004)	5	3	2	3	1	4	4	2	3	4	2	1
Philadelphia (1993)	4	4	2	3	1	3	4	3	2	4	1	2
Amelie (2001)		2					2					
Malena (2000)		2					1					
read review		2					2					
Goodbye Lenin! (2004)		2					2					
Vidocq		1					1					
28 Days Later (2003)	3	3	1	4	2	4	3	1	1	5	2	1
Resident Evil (2002)	3	4	1	4	1	4	3	1	2	4	2	1
Sleepy Hollow (1999)	3	4	1	4	1	3	4	2	2	5	1	1
The Ring (2002)	4	3	2	5	1	3	4	1	1	4	1	1
Casablanca	1	3	3	1	4	2	2	4	3	1	4	5
Gone With the Wind	2	2	3	2	4	3	3	4	3	2	5	4
West Side Story	1	3	3	2	4	3	3	4	3	1	5	5
Roman Holiday	2	2	3	1	5	2	3	4	4	1	5	4
An Affair to Remember	1	3	4	2	4	3	2	3	3	1	5	4
The Way We Were	1	2	3	2	5	2	2	3	4	2	4	5
Doctor Zhivago	2	2	3	2	5	2	3	3	3	1	5	5
It's a Wonderful Life	1	2	3	1	4	3	2	4	3	1	5	4
Love Story	2	3	4	2	5	2	3	3	4	1	4	4
City Lights.	1	3	4	1	4	2	2	4	3	1	5	5
MÉDIA DAS AVALIAÇÕES	3,11	3,12	2,95	2,84	3,00	3,27	3,12	3,00	3,11	2,95	3,16	3,02

Filtragem colaborativa

- Matriz de correlação entre usuários

[illegible]

Filtragem colaborativa

- Matriz de distancia entre usuários
 - Distância = $1 - \text{correlação}$

	Maria	José	João	Pedro	Laura	Manoel	Jaime	Rafael	Lucas	Reinaldo	Felipe	Eva
Maria	0,00	0,41	1,40	0,32	1,70	0,45	0,33	1,52	1,41	0,28	1,75	1,73
José	0,41	0,00	1,46	0,53	1,35	0,36	0,24	1,51	1,44	0,45	1,45	1,34
João	1,40	1,46	0,00	1,51	0,40	1,45	1,56	0,23	0,23	1,58	0,42	0,41
Pedro	0,32	0,53	1,51	0,00	1,63	0,38	0,34	1,70	1,49	0,22	1,66	1,67
Laura	1,70	1,35	0,40	1,63	0,00	1,34	1,41	0,44	0,43	1,66	0,12	0,13
Manoel	0,45	0,36	1,45	0,38	1,34	0,00	0,34	1,54	1,40	0,41	1,29	1,36
Jaime	0,33	0,24	1,56	0,34	1,41	0,34	0,00	1,64	1,45	0,44	1,45	1,47
Rafael	1,52	1,51	0,23	1,70	0,44	1,54	1,64	0,00	0,28	1,63	0,44	0,37
Lucas	1,41	1,44	0,23	1,49	0,43	1,40	1,45	0,28	0,00	1,46	0,42	0,44
Reinaldo	0,28	0,45	1,58	0,22	1,66	0,41	0,44	1,63	1,46	0,00	1,73	1,68
Felipe	1,75	1,45	0,42	1,66	0,12	1,29	1,45	0,44	0,42	1,73	0,00	0,15
Eva	1,73	1,34	0,41	1,67	0,13	1,36	1,47	0,37	0,44	1,68	0,15	0,00

Filtragem colaborativa

- Seleção dos “3” mais próximos
 - A cor salmão indica os 3 mais próximos de Jose

	Maria	José	João	Pedro	Laura	Manoel	Jaime	Rafael	Lucas	Reinaldo	Felipe	Eva
Maria	0,00	0,41	1,40	0,32	1,70	0,45	0,33	1,52	1,41	0,28	1,75	1,73
José	0,41	0,00	1,46	0,53	1,35	0,36	0,24	1,51	1,44	0,45	1,45	1,34
João	1,40	1,46	0,00	1,51	0,40	1,45	1,56	0,23	0,23	1,58	0,42	0,41
Pedro	0,32	0,53	1,51	0,00	1,63	0,38	0,34	1,70	1,49	0,22	1,66	1,67
Laura	1,70	1,35	0,40	1,63	0,00	1,34	1,41	0,44	0,43	1,66	0,12	0,13
Manoel	0,45	0,36	1,45	0,38	1,34	0,00	0,34	1,54	1,40	0,41	1,29	1,36
Jaime	0,33	0,24	1,56	0,34	1,41	0,34	0,00	1,64	1,45	0,44	1,45	1,47
Rafael	1,52	1,51	0,23	1,70	0,44	1,54	1,64	0,00	0,28	1,63	0,44	0,37
Lucas	1,41	1,44	0,23	1,49	0,43	1,40	1,45	0,28	0,00	1,46	0,42	0,44
Reinaldo	0,28	0,45	1,58	0,22	1,66	0,41	0,44	1,63	1,46	0,00	1,73	1,68
Felipe	1,75	1,45	0,42	1,66	0,12	1,29	1,45	0,44	0,42	1,73	0,00	0,15
Eva	1,73	1,34	0,41	1,67	0,13	1,36	1,47	0,37	0,44	1,68	0,15	0,00

Filtragem colaborativa

- Previsão de avaliação
 - Assuma que entrou mais um filme: “Os Normais”
- Qual deve ser a avaliação de José (alvo)?
 - Jaime avaliou “Os Normais” como “2”
 - Manoel avaliou “Os Normais” como “3”
 - Maria avaliou “Os Normais” como “4”

$$P_{u,i} = \bar{r}_u + \frac{\sum_{j=1}^n w_{u,j} (r_{j,i} - \bar{r}_j)}{\sum_{j=1}^n |w_{u,j}|}$$

	Avaliação para "Os Normais"	Avaliação Média dos filmes pelo usuário	Dif.entre avaliação do filme alvo para avaliação média	Pearson entre José e usuário similar	
	$r_{j,i}$	$\bar{r}_u$	$(r_{j,i} - \bar{r}_j)$	$w_{u,j}$	$w_{u,j}(r_{j,i} - \bar{r}_j)$
Jaime	2	3,1	-1,1	0,76	-0,836
Manoel	3	3,3	-0,3	0,64	-0,192
Maria	4	3,1	0,9	0,59	0,531
José	???	3,1		1,99	-0,497

$$P_{José, "Os Normais"} = 3,1 + \frac{(-0,497)}{1,99} = 2,85$$

$$\sum_{j=1}^n |w_{u,j}| \quad \sum_{j=1}^n w_{u,j} (r_{j,i} - \bar{r}_j)$$

Filtragem colaborativa

- **Vantagens**

- Extrema eficiência e poder discriminatório
- Alta relevância das recomendações
- Recomendações melhoram à medida que aumenta o número de usuários ou aumenta o número de transações no Banco de Dados

- **Desvantagens**

- Consumir muito tempo e recursos computacionais → $O(n^2)$
- Partida a frio → não funciona sem matriz preenchida
- Esparsidade da matriz
- Não atende usuários novatos que não manifestaram suas preferências
- Itens não avaliados (lançamentos) ficaram fora da recomendação
 - Tendência ao popularesco

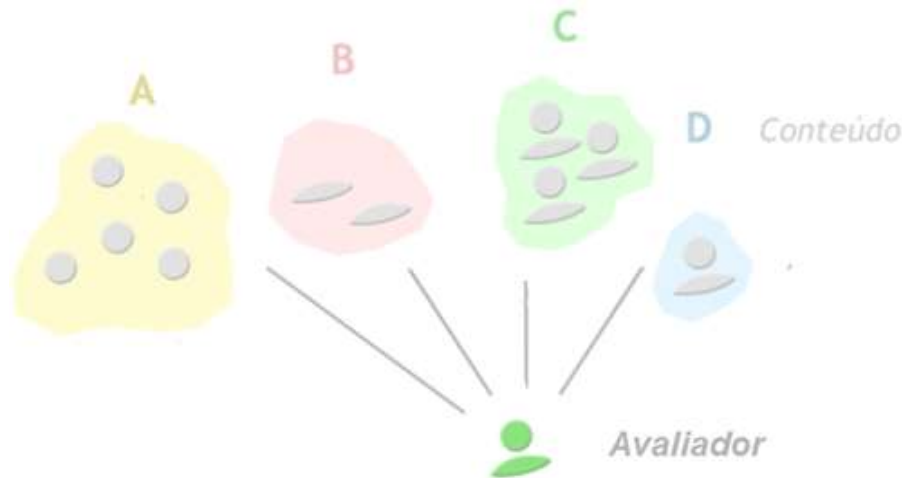
RECOMENDAÇÃO BASEADA EM CONTEÚDO

Recomendação baseada em conteúdo

- Recomendações baseadas em informações sobre o conteúdo dos itens em lugar da opinião de outros usuários.
- Usa algoritmos de aprendizagem de máquinas para induzir um perfil das preferências de um usuário a partir de exemplos tendo em vista uma descrição de características de conteúdo (Agrupamentos).
- Existem muitas aplicações usuais que utilizam este método.

Recomendação baseada em conteúdo

- Método
 - Passo 1: classifique os itens avaliados segundo uma categoria
 - Passo 2: para cada categoria, calcule a avaliação média de cada avaliador
 - Passo 3: ordene para obter preferências por categorias
 - Passo 4: calcule a previsão de avaliação



Recomendação baseada em conteúdo

- Para calcular a previsão de avaliação do usuário alvo a um produto não conhecido por ele

$$P_{u,i} = \frac{\bar{r}_i * \bar{r}_{categoria_i}}{\bar{r}_{categoria}}$$

$\bar{r}_{categoria}$ → “Avaliação_Média_da_Categoria”

$\bar{r}_i$ → “Avaliação_Média_do_Itemi”

$\bar{r}_{categoria_i}$ → “Avaliação_Média_do_usuario_U
da_Categoria_da_qual_o_item_pertence”

Recomendação baseada em conteúdo

- Usando o mesmo exemplo, mas com os filmes categorizados.

Ação	Maria	José	João	Pedro	Laura	Manoel	Jaime	Rafael	Lucas	Reinaldo	Felipe	Eva
Gladiator (2000)	4	5	2	3	3	4	4	3	3	4	3	4
Braveheart (1995)	4	4	2	4	3	4	5	2	3	4	3	4
Hannibal (2001)	3	5	2	3	3	5	5	2	3	3	3	3
Black Hawk Down (2001)	4	4	3	3	3	5	4	2	3	4	3	3
The Matrix (1999)	4	4	3	4	3	4	4	2	3	4	3	4
Pulp Fiction (1994)	3	4	3	3	4	5	4	3	3	4	4	4
Training Day (2001)	4	4	3	3	4	4	4	3	3	4	4	3
Gangs of New York (2002)	4	4	3	4	4	5	5	2	3	4	4	3
Pearl Harbor (2001)	4	5	3	4	4	4	5	2	3	3	4	3
Spawn (1997)	3	4	2	3	3	4	4	3	2	3	3	3
	3,70	4,30	2,60	3,40	3,40	4,40	4,40	2,40	2,90	3,70	3,40	3,40

Recomendação baseada em conteúdo

Comédia	Maria	José	João	Pedro	Laura	Manoel	Jaime	Rafael	Lucas	Reinaldo	Felipe	Eva
The Original Kings Comedy (2000)	2	3	4	2	4	2	2	5	5	3	3	4
Ice Age (2002)	3	2	5	3	3	3	3	5	5	2	4	3
Finding Nemo (2003)	3	3	4	2	3	2	3	4	5	2	4	4
Zoolander (2001)	3	3	4	2	4	3	2	5	4	3	3	4
Kangaroo Jack (2003)	2	2	4	3	4	3	2	4	5	3	4	4
Bruce Almighty (2003)	2	2	4	3	3	3	2	4	4	3	4	4
Scary Movie (2000)	2	3	4	2	4	3	3	4	5	2	4	3
Bad Boys II (2003)	2	2	5	3	4	3	2	4	4	2	4	4
Men in Black (1997)	3	3	5	2	4	2	2	5	4	3	4	4
	2,44	2,56	4,33	2,44	3,67	2,67	2,33	4,44	4,56	2,56	3,78	3,78

Recomendação baseada em conteúdo

Dramas	Maria	José	João	Pedro	Laura	Manoel	Jaime	Rafael	Lucas	Reinaldo	Felipe	Eva
Nixon (1995)	5	4	3	3	2	3	4	3	2	3	1	1
The Pianist (2002)	4	3	2	4	2	3	4	2	3	4	2	1
Hamlet (1996)	5	4	3	3	2	3	3	2	3	4	2	1
Seven (1995)	4	3	2	3	1	3	4	2	2	4	1	1
Malena (2000)	5	4	3	3	1	4	3	3	3	3	2	2
read review	5	3	2	4	1	4	4	3	3	3	2	1
Surviving Picasso (1996)	5	4	2	3	2	4	3	2	2	4	2	2
A Few Good Men (1992)	4	3	3	3	1	3	4	2	3	4	2	2
The Mothman Prophecies (2002)	5	4	3	4	2	4	4	3	2	3	1	2
Goodbye Lenin! (2004)	5	3	2	3	1	4	4	2	3	4	2	1
Philadelphia (1993)	4	4	2	3	1	3	4	3	2	4	1	2
	4,64	3,55	2,45	3,27	1,45	3,45	3,73	2,45	2,55	3,64	1,64	1,45

Recomendação baseada em conteúdo

Arte	Maria	José	João	Pedro	Laura	Manoel	Jaime	Rafael	Lucas	Reinaldo	Felipe	Eva
Amelie (2001)		2					2					
Malena (2000)		2					1					
read review		2					2					
Goodbye Lenin! (2004)		2					2					
Vidocq		1					1					
		1,80					1,60					

Horror	Maria	José	João	Pedro	Laura	Manoel	Jaime	Rafael	Lucas	Reinaldo	Felipe	Eva
28 Days Later (2003)	3	3	1	4	2	4	3	1	1	5	2	1
Resident Evil (2002)	3	4	1	4	1	4	3	1	2	4	2	1
Sleepy Hollow (1999)	3	4	1	4	1	3	4	2	2	5	1	1
The Ring (2002)	4	3	2	5	1	3	4	1	1	4	1	1
	3,25	3,50	1,25	4,25	1,25	3,50	3,50	1,25	1,50	4,50	1,50	1,00

Recomendação baseada em conteúdo

Romance	Maria	José	João	Pedro	Laura	Manoel	Jaime	Rafael	Lucas	Reinaldo	Felipe	Eva
Casablanca	1	3	3	1	4	2	2	4	3	1	4	5
Gone With the Wind	2	2	3	2	4	3	3	4	3	2	5	4
West Side Story	1	3	3	2	4	3	3	4	3	1	5	5
Roman Holiday	2	2	3	1	5	2	3	4	4	1	5	4
An Affair to Remember	1	3	4	2	4	3	2	3	3	1	5	4
The Way We Were	1	2	3	2	5	2	2	3	4	2	4	5
Doctor Zhivago	2	2	3	2	5	2	3	3	3	1	5	5
It's a Wonderful Life	1	2	3	1	4	3	2	4	3	1	5	4
Love Story	2	3	4	2	5	2	3	3	4	1	4	4
City Lights.	1	3	4	1	4	2	2	4	3	1	5	5
	1,40	2,50	3,30	1,60	4,40	2,40	2,50	3,60	3,30	1,20	4,70	4,50

Recomendação baseada em conteúdo

- Média de avaliações por categoria, por usuário

Usuário	Comédia	Romance	Ação	Arte	Drama	Horror
Maria	2,44	1,4	3,7		4,64	3,25
José	2,56	2,5	4,3	1,8	3,55	3,5
João	4,33	3,3	2,6		2,45	1,25
Pedro	2,44	1,6	3,4		3,27	4,25
Laura	3,67	4,4	3,4		1,45	1,25
Manoel	2,67	2,4	4,4		3,45	3,5
Jaime	2,33	2,5	4,4	1,6	3,73	3,5
Rafael	4,44	3,6	2,4		2,45	1,25
Lucas	4,56	3,3	2,9		2,55	1,5
Reinaldo	2,56	1,2	3,7		3,64	4,5
Felipe	3,78	4,7	3,4		1,64	1,5
Eva	3,78	4,5	3,4		1,45	1

- Média de avaliações por categoria

Comédia	3,30
Romance	2,95
Ação	3,50
Arte	1,70
Drama	2,86
Horror	2,52

Recomendação baseada em conteúdo

Novos Filmes	Classe	Pedro	Laura	Manoel	Jaime	Rafael	Lucas	Reinaldo	Felipe	Eva	Avaliação Média do Filme
Os Normais	comédia	2	4	3	2	4	4	2	4	4	3,22
Love Story	romance	2	5	2	3	3	4	1	4	4	3,11
Hannibal (2001)	ação	3	2	4	5	2	3	3	2	1	2,78
Gangs of New York (2002)	ação	4	1	4	4	3	3	3	1	2	2,78
The Pianist (2002)	drama	4	2	3	4	2	3	4	2	1	2,78
Sleepy Hollow (1999)	horror	3	1	3	4	2	2	5	1	1	2,44

Novos Filmes	Classe	Avaliação Média da Categoria	Avaliação Média do Filme	Aval média do José por Categoria	Avaliação José para Filme
Os Normais	comédia	3,30	3,22	2,56	2,50
Love Story	romance	2,95	3,11	2,5	2,64
Hannibal (2001)	ação	3,50	2,78	4,3	3,42
Gangs of New York (2002)	ação	1,70	2,78	1,8	2,94
The Pianist (2002)	drama	2,86	2,78	3,55	3,46
Sleepy Hollow (1999)	horror	2,52	2,44	3,5	3,39

Recomendação baseada em conteúdo

- Vantagens
 - Não necessita de dados de outros usuários.
 - Sem problemas de Partida a Frio ou de Esparsidade.
 - Permite recomendações a usuários com gostos exclusivos
 - Capaz de recomendar itens novos ou não populares
 - Sem problema de Primeira Avaliação.
 - Pode fornecer explicações sobre os itens recomendados listando as características do conteúdo que conduziram à recomendação.
 - Tecnologia madura
- Desvantagens
 - Recomendações serem estáticas, desprezando a opinião dos usuários
 - O conteúdo precisa ser codificado por características inteligíveis.
 - Baixa eficiência se o conteúdo não for muito informativo
 - Para livros técnicos o título é relevante mas para novelas ou filmes o título pode trazer pouca informação

DESCOBERTA: REGRAS DE ASSOCIAÇÃO

Regras de associação

- Descoberta de Conhecimento
 - Regra de associação (Agrawal 1993)
 - Cada tupla corresponde a uma transação e consiste conjunto de itens
- Regra de associação $X \Rightarrow Y$
 - X e Y são conjunto de itens, $X \cap Y = \emptyset$
 - Fator de suporte F_{sup} e um fator de confiança F_{conf}
 - Suporte refere-se a frequência com que a $X \wedge Y$ aparecem juntos no banco
 - N é o número total de tuplas no banco
 - Confiança representa a chance de aparecer Y uma vez que temos X

Regras de associação

- Exemplo, suporte=30% e confiança=70%

ID	Hannibal	Bad Boys	Os Normais	A pequenina	Batman	A Era do Gelo III	Up
1	😊	😊		😊			
2	😊	☠️	😊	☠️	😊	☠️	
3	☠️	😊		😊			
4	😊		😊	☠️	☠️	😊	😊
5	☠️	😊		☠️			
6		😊	☠️	😊	😊		
7	😊	☠️		☠️	☠️	😊	😊
8		😊	☠️	☠️	😊		
9	😊	☠️			☠️	😊	😊
10	☠️	😊		😊	☠️		

$$Fsup_{BadBoys,Pequenina} = \frac{4}{10} = 40\%$$

$$Fconfiança_{BadBoys \rightarrow Pequena} = \frac{40\%}{40\%} = 100\%$$

$$Fconfiança_{Pequenina \rightarrow BadBoys} = \frac{40\%}{60\%} = 66\%$$

Regras de associação

- Agrupamentos
 - Recomendação pelo gosto do grupo
- Filtragem colaborativa
 - Recomendação por correlação entre usuário
- Regras de associação
 - Recomendação pela correlação entre itens

CLASSIFICADORES

Classificadores

- Classificadores são modelos computacionais gerais
 - Suas entradas são
 - As características (ou atributos) dos itens
 - As preferências dos usuários
 - O relacionamento entre os itens
 - Suas saída são
 - Uma classificação (divisão em classes)
 - Uma hierarquização das classes
 - Uma estimativa de preferência.
- Classificadores adquirem conhecimentos através de conjuntos de treinamento
- Métodos ou procedimentos classificadores usuais são:
 - As redes neurais
 - As redes Bayesianas
 - O modelo de indução de regras
 - etc...

Classificadores

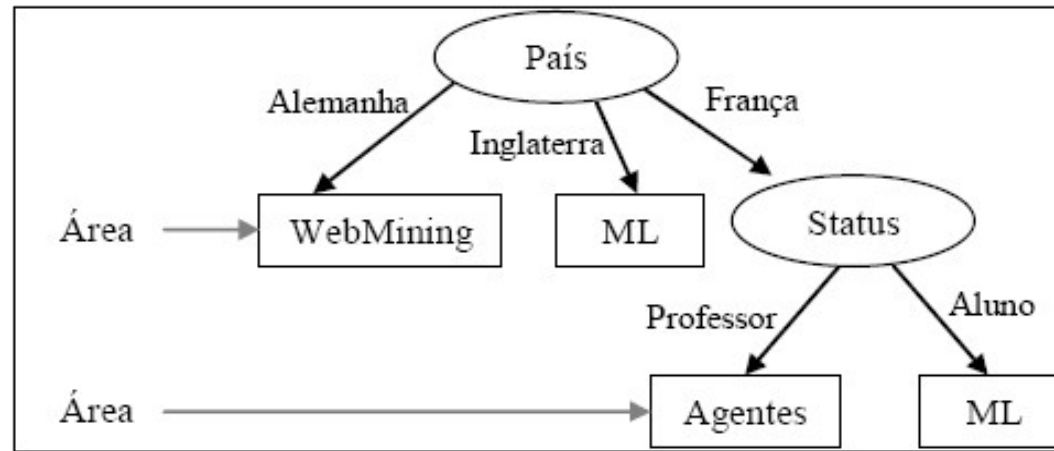
- Classificadores → Aprendizado supervisionado
 - Consiste em descobrir algum relacionamento entre os atributos preditivos e o atributo objetivo, para revelar um conhecimento que possa ser utilizado para prever classes de uma tupla desconhecida, ou seja, que não possui uma classificação.
- As tuplas consistem de atributos preditivos e um atributo objetivo, o último indicando a qual classe a tupla pertence
- Técnicas
 - Redes Neurais
 - Regras de Decisão

Classificadores

- Exemplo
 - Atributo objetivo em cinza
 - Árvore de Decisão [Cabena 1997].
 - Conhecimento descoberto apresentado frequentemente por Regras
 - Atributo objetivo → área de interesse

Função	Pais	Área de Interesse
Professor	França	Agentes
Pesquisador	Inglaterra	ML
Aluno	França	ML
Pesquisador	Inglaterra	ML
Professor	França	Agentes
Aluno	Alemanha	WebMining
Aluno	Alemanha	WebMining
Pesquisador	Alemanha	ML
Professor	França	WebMining
Professor	França	ML
Aluno	França	ML

Classificadores



- Regras obtidas no exemplo:
 - Se (PAÍS = Alemanha) então Área = WebMining
 - Se (PAÍS = Inglaterra) então Área = ML
 - Se (PAÍS = França e Status = Professor) então Área = Agentes
 - Se (PAÍS = França e Status = Aluno) então Área = ML
- Fácil para um sistema de recomendação decidir o que recomendar no caso de um novo usuário cadastrado.

Classificadores

- Vantagens
 - Versatilidade
 - Possibilidade de combinação com outros métodos para aumentar a precisão das recomendações
- Desvantagem
 - Necessidade de conjuntos de treinamento relevantes

AGRUPAMENTO

Agrupamento

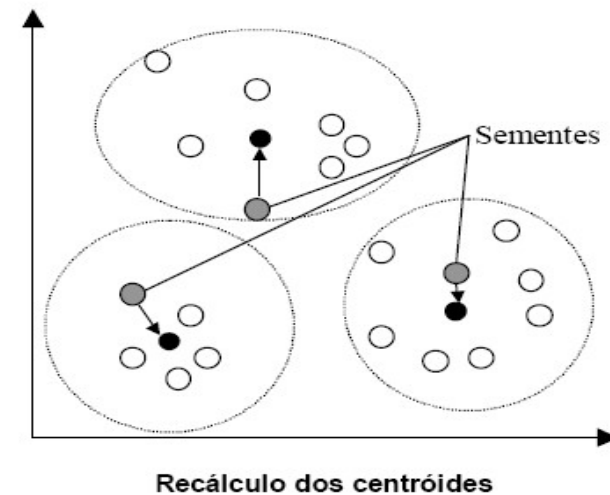
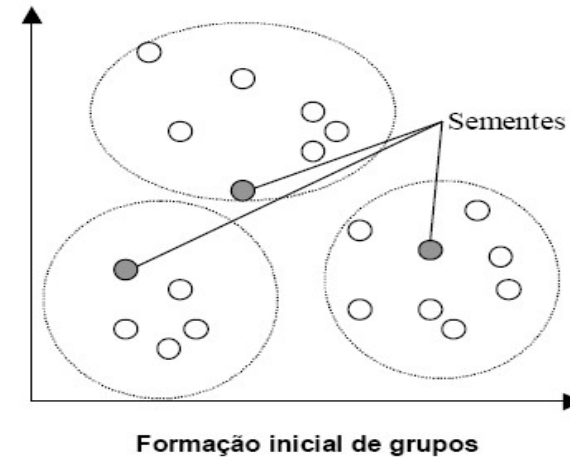
- Aprendizado não supervisionado
 - Usuários classificados em grupos de perfis semelhantes
 - Preferências dos grupos são calculadas
 - Recomendação é feita baseada na preferência do grupo
- Grupos são partições de um banco de dados em conjunto de tuplas similares
- agrupamento de qualidade similaridade intra-classe é alta e a inter-classe é baixa

Agrupamento

- Usuários classificados em grupos de perfis semelhantes
- Preferências dos grupos são calculadas
- Recomendação é feita baseada na preferência do grupo
- Diferenciação
 - Agrupamentos
 - Número fixo de agrupamentos para todo o Banco de Dados
 - Filtragem colaborativa
 - Para cada usuário, determina-se seu particular agrupamento de usuários de perfil semelhante

Agrupamento

- Técnica K-means
 - Passo 1: encontrar k casos para servirem de “semente” para o processo (os k primeiros casos)
 - Passo 2: Para cada um dos registros armazenados no BD, verificar a semente mais próxima dele
 - Passo 3: determinar os novos centróides a partir de todos os elementos existentes em cada agrupamento
 - Retorne ao passo 3 até que o centróide permaneça o mesmo ou por limitação de recursos



Agrupamento

- Técnica de agrupamentos
- Vantagens
 - Mais rápida do que a filtragem colaborativa, pois os agrupamentos são genéricos e não personalizados
 - Pode-se usar esta técnica como um passo inicial para restringir o número de usuários selecionado para um algoritmo de filtragem colaborativa
- Desvantagens
 - As recomendações por agrupamento são menos relevantes de que as recomendações individuais geradas pela filtragem colaborativa

PARTE 4: INDICAÇÕES DE MODELOS DE SISTEMAS DE RECOMENDAÇÃO

Projeto de um SR

- Definição do objetivo do negócio indica
 - Método
 - Tipo de entrega
 - Grau de personalização
- Características da base de produtos
 - Tamanho
 - Frequência de atualização

Objetivos de negócios e modelos de sistemas de recomendação

- Objetivos de negócios indicam modelos de aplicações de sistemas de recomendação correspondentes.
- Alguns Modelos de Aplicações são:
 - Listas de recomendação genéricas
 - Listas de recomendação baseadas em buscas
 - Avaliações e comentários de consumidores
 - Serviços de notificação
 - Recomendações associadas a produtos
 - Personalização profunda

Listas de recomendação genericas

- Objetivo de negócios: atender a visitantes novos ou pouco frequentes
- Características: usualmente o usuário recebe uma lista de hyperlinks para ter acesso a recomendações. O sistema requer quase nenhuma informação do visitante
- Tipo de entrega: sob demanda ou “pull”
- Grau de personalização: sugestões não personalizadas

Listas de recomendação baseadas em buscas

- Objetivo de negócios: atender a usuários com necessidades definidas
- Características: o usuário fornece uma palavra chave ou atributo para ter acesso a recomendações.
- Tipo de entrega: sob demanda ou “pull”
- Grau de personalização: personalização efêmera de produtos ou serviços

Avaliações e comentários de consumidores

- Objetivo de negócios: obter credibilidade através da opinião da comunidade
- Características: vendas a varejo, principalmente no comércio eletrônico necessitam exorcizar o fantasma da baixa credibilidade. Sistemas de Recomendação, podem utilizar a navegação do usuário alvo para sugerir a ele a obtenção de revisões não personalizadas, comentários, avaliações e previsões. Esta opinião da comunidade é mais bem aceita pelos usuários
- Tipo de entrega: entrega passiva
- Grau de personalização: não personalizadas

Serviços de notificação

- Objetivo de negócios: trazer de volta antigos clientes
- Características: o comércio que conhece os interesses de seus clientes pode usar esta informação para trazê-los de volta notificando-os por ocasião de lançamentos ou promoções de produtos de seu interesse. Utilizando palavras chave obtidas do cliente alvo e atributos de itens sendo recomendados faz-se recomendações de produtos ou serviços
- Tipo de entrega: entrega pró-ativa (“push”)
- Grau de personalização: sugestões personalizadas persistentes buscando construir relações mais estreitas com a clientela

Recomendações associadas a produtos

- Objetivo de negócios: fazer vendas “casadas”
- Características: utilizando a navegação corrente do consumidor alvo como uma indicação de interesse
Sistemas deste tipo empregam correlação item-a-item e o histórico de compras da comunidade. A escolha do que recomendar é um caso clássico de Mineração de Dados
- Tipo de entrega: entrega na forma passiva
- Grau de personalização: personalização efêmera de produtos ou serviços

Personalização profunda

- Objetivo de negócios: construção de relacionamentos duradouros
- Características: clientes fiéis são garantia de vendas por isso tem
 - Preferências conhecidas
 - Tratamento especial
 - Relacionamento vantajoso para ambos

Baseia-se no histórico de preferências, aquisições e navegação. O vínculo é o mais forte porém, é de difícil implementação. Usa-se o histórico do consumidor alvo e a filtragem colaborativa

- Tipo de entrega: sob demanda ou “pull”
- Grau de personalização: personalização persistente

Influência do domínio de produtos

- O domínio influi e para verificar a necessidade de sistemas de recomendações é necessário considerar:
 - O número de itens em um site
 - Itens de elevado preço.
 - Intervalo de tempo entre atualizações de catálogo
 - Homogeneidade de produtos no domínio
 - Proporção de produtos relacionados por atributos específicos.

O número de itens em um site

- Quando este número é pequeno, quase não há necessidade de Sistemas de Recomendação

Itens de elevado preço

- Quando a aquisição de um item for algo que exija um estudo financeiro é interessante dispor de Sistemas de Recomendação

Intervalo de tempo entre atualizações de catálogo

- Catálogos estáveis dão tranquilidade a usuários enquanto catálogos voláteis requerem o uso de Sistemas de Recomendação

Homogeneidade de produtos no domínio

- Quando a variedade de itens semelhantes é muito grande os Sistemas de Recomendação são ajuda valiosa

Proporção de produtos relacionados por atributos específicos

- A seleção por atributos específicos exige pouco dos Sistemas de Recomendação que são tão mais úteis quanto menos o consumidor tiver certeza do que quer adquirir

PARTE 5: CONCLUSÃO

- Conceitos Fundamentais de um SR
- Caracterização de SR pela entrada e saída
- Caracterização pelos itens do projeto
- Caracterização pela técnica de geração da recomendação
- Projeto de um SR

Referências

- R Schafer, J.B., Konstan, J., and Riedl, J., (2001) Electronic Commerce Recommender Applications. Journal of Data Mining and Knowledge Discovery, vol. 5 nos. 1/2, pp. 115-152
- www.cs.uni.edu/~schafer/courses/189/thesis_ch3.pdf
- <http://www.ee.ucl.ac.uk/~icox/course/presentations/RecommenderSystems.ppt>