

Capítulo

15

Sistemas de Recomendação

Claudia Motta, Ana Cristina Bicharra, Adriana Vivacqua, Flávia Maria Santoro, Jonice Oliveira

Meta. *Apresentar conceitos relacionados ao desenvolvimento de sistemas de recomendação.*

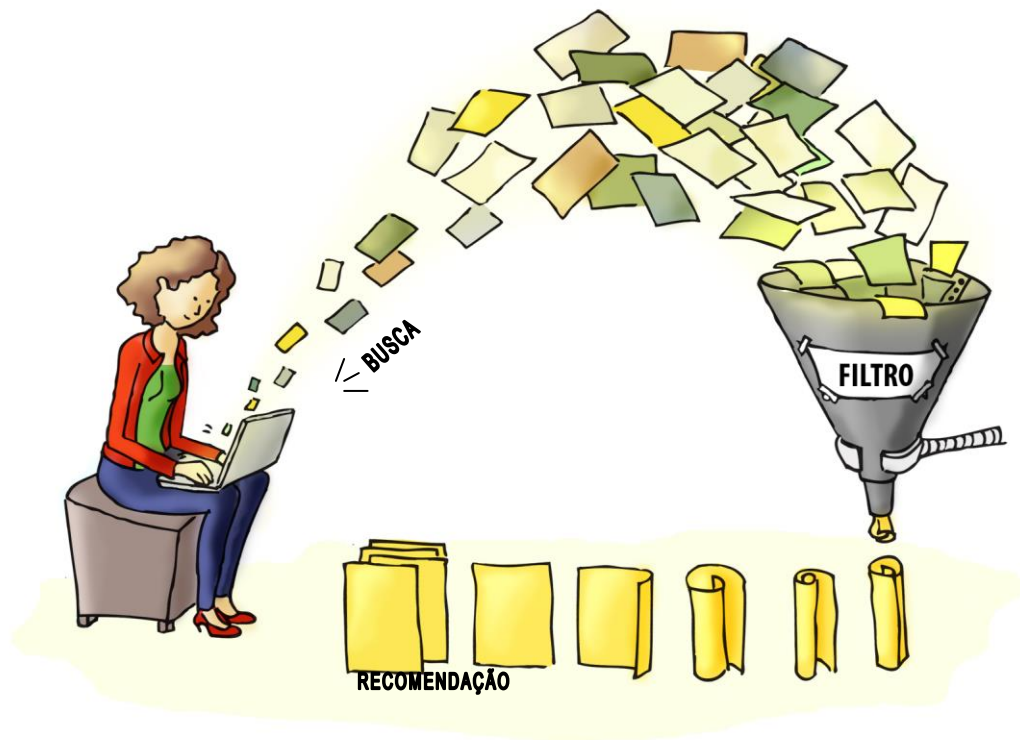
Objetivos educacionais. *Após o estudo desse capítulo, o aluno deve ser capaz de:*

- *Listar os diferentes tipos de sistemas de recomendação;*
- *Relacionar as características e requisitos para o desenvolvimento de um sistema de recomendação;*
- *Aplicar métodos utilizados em sistemas de recomendação que atuem sobre uma base de dados.*

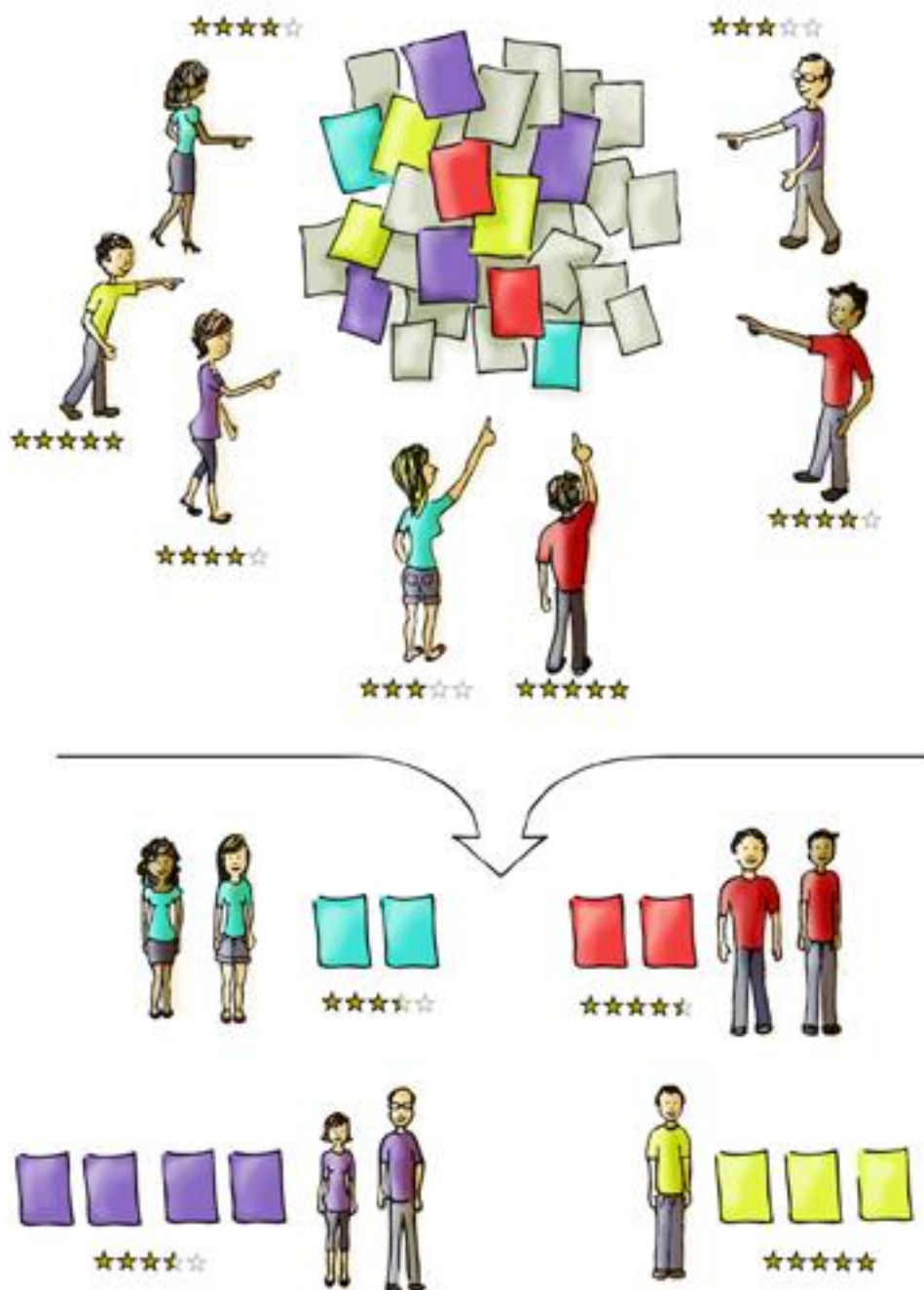
Resumo. *Sistemas de recomendação visam oferecer, a partir de grandes volumes de informações, aquilo que pode interessar especificamente ao usuário. As técnicas, em geral, dependem das contribuições dos indivíduos na avaliação da informação. O princípio dos sistemas de recomendação se baseia em “o que é relevante para mim, também pode ser relevante para alguém com interesse similar”. Neste capítulo são apresentados os conceitos fundamentais de sistemas de recomendação, que vão desde a entrada e saída esperadas, grau de atuação e personalização da recomendação, até a utilização das técnicas de geração de recomendações.*

15.1 O que são e como funcionam os Sistemas de Recomendação

Vivemos num mundo inundado de informações que aumentam numa enorme velocidade graças aos avanços nas tecnologias de informação e comunicação (TICs). Somos todos contribuintes e consumidores de informação. É um desafio ter acesso às informações que nos apoiem nas decisões corretas. Técnicas de recuperação de informação auxiliam. A premissa básica é que sabemos o que procuramos. O usuário menciona um conjunto de palavras ou indicadores e os documentos identificados como os mais relevantes são recomendados por algoritmos de busca e recuperação.



Sistemas de recomendação são filtros de informação para apresentar itens ou objetos – como páginas web, filmes, músicas, livros, medicamentos, lojas, artigos – que provavelmente são do interesse do usuário [Shaffer 2001]. O princípio dos sistemas de recomendação se baseia em “o que é relevante para mim, também pode ser relevante para alguém com interesse similar”. A grande maioria das técnicas de recomendação depende da avaliação da informação realizada pelos indivíduos.



Podemos até não notar, mas sistema de recomendação já é uma realidade embutida em muitos sites de compra na Internet. Todos nós já nos deparamos com sugestões intercaladas às informações que acessamos na rede, como pode ser visto quando fazemos compras na Amazon (*"Today's recommendations for you"*).

Os principais componentes de um Sistema de Recomendação são cliente e produto. Um produto é um recurso que pode ser de diferentes naturezas, por exemplo: um conteúdo, um arquivo, uma informação, uma pessoa, um objeto. A recomendação é uma função de mapeamento de interesses do cliente para obtenção de um ou mais produtos.

Uma recomendação R para a escolha de “ p ” deve ser feita de tal forma que

$$R(c,p) = \max F(c, p_i)$$

Considere que:

“ c ” representa o cliente que usa o Sistema de Recomendação

P é um conjunto de produtos disponíveis à avaliação

$$p_i \in P$$

F é a função que determina a relevância de p_i em relação a c .

Todo o processo de sistemas de recomendação gira em torno da definição dessa função que avalia a utilidade de um produto para um certo cliente. Chamamos o usuário de um Sistema de Recomendação de cliente para ressaltar a sua importância no processo. Geralmente, F leva em consideração a similaridade entre perfis dos clientes.

História dos Sistemas de Recomendação

Os precursores de Sistema de Recomendação se encontram em trabalhos de recuperação de informação, teorias de aproximação, ciência cognitiva e teorias de previsão. No entanto, a área de recomendações começou a se estabelecer na década de 1990, quando pesquisadores passaram a enfocar o problema da recomendação em si. Um dos primeiros sistemas de recomendação foi o RINGO (Shardanand e Maes, 1995), que recomendava músicas a partir do perfil do usuário gerado por informações explicitamente fornecidas pelos usuários. O sistema se ajustava continuamente com o uso prolongado e sucessivas recomendações. Grouplens (Resnick *et al.*, 1994), criado na mesma época, já propunha uma arquitetura genérica para recomendação de notícias. Este sistema mais tarde evoluiu para o MovieLens (Konstan, 1997) em que sugestões de filmes eram geradas a partir de correlações entre avaliações dos usuários sem utilizar características predefinidas para descrição do usuário. Estes trabalhos iniciais ajudaram a impulsionar e solidificar a área de Sistema de Recomendação.

A tecnologia usada no sistema RINGO transformou-se em um produto chamado Firefly, um site de recomendação de músicas e artistas, livros e comunidades. Por meio de parcerias com diversas empresas, a Firefly difundiu o sistema de recomendação e a filtragem colaborativa no meio organizacional. Algumas grandes empresas adotaram essa tecnologia, como Yahoo!, ZDNet e Barnes and Noble.

Grandes companhias como Amazon.com e eBay utilizam Sistemas de Recomendação. O sistema da Amazon, por exemplo, disponibiliza várias sugestões baseadas em técnicas distintas; o eBay possibilita o “direito de resposta” aos donos de produtos sobre avaliações que lhes foram atribuídas.

Os Sistemas de Recomendação são caracterizados de acordo com três eixos:

Tipo de Entrada e Saída

Itens do Projeto

Método de Recomendação

Um Sistema de Recomendação é um sistema colaborativo porque a recomendação é feita a partir da organização, manipulação, sumarização e agrupamento das avaliações individuais. O indivíduo contribui para avaliação de produtos que serão

consumidos por outros indivíduos. É um processo coletivo, embora a interação seja assíncrona e os benefícios não sejam percebidos no momento da contribuição individual ao grupo.

Combinação Social

Existe um tipo particular de recomendação, que é o de recomendar pessoas, denominado “combinação social”. Pense na senhorinha casamenteira de antigamente: sempre a procurar um par ideal para cada moça ou rapaz solteiro da cidade. Esse também é o objetivo da combinação social. Os sistemas baseiam-se em perfis de usuário para encontrar e sugerir a interação entre dois usuários.

Quando você não sabe resolver um problema, o que você faz? Você procura um amigo ou conhecido que possa lhe ajudar. Este mesmo princípio se aplica a sistemas de recomendação. Se você não sabe a quem procurar, o sistema pode lhe sugerir um nome. Quem é o especialista? Quem tem a habilidade necessária ou os recursos desejados? Quem está no nível de conhecimento correto, tem personalidade compatível, está no local certo e na hora certa? Os critérios a serem usados para a combinação social são os mais variados, mas em geral as recomendações são baseadas em um mapeamento de características ou habilidades dos usuários. Por isso geralmente é necessário montar um perfil de usuário, um mapa que descreva as competências e habilidades do usuário com o detalhe necessário, e definir as necessidades de outros usuários. Assim, quando um usuário pergunta “quem tem capacidade de liderança?”, o sistema confere os perfis cadastrados e determina os candidatos que o usuário deve contatar.

15.2 Fontes de entrada e saída para recomendações

Para gerar recomendações é preciso considerar a fonte dos dados. Quem provê as informações que servirão de base para a recomendação? As informações podem ser oriundas do consumidor alvo ou de dados gerados pela comunidade. Também temos que considerar se a recomendação será para um cliente específico ou para uma comunidade em geral.

É sempre uma opção válida perguntar diretamente ao cliente suas preferências e o que ele está buscando. Porém, se o usuário já sabe o que quer, uma intervenção ou sugestão pode atrapalhar ou retardar a obtenção de seu objetivo. Por isso algumas informações sobre os clientes são obtidas de forma indireta.

A navegação explícita pelas páginas da Internet, registrada no histórico de páginas visitadas, é uma indicação do interesse do usuário. Essa informação pode dar a um Sistema de Recomendação pistas sobre os gostos do consumidor alvo e sobre a sua avaliação do conteúdo das páginas. Também podemos lançar mão do histórico de consultas prévias realizadas extraindo palavras-chave da consulta ou mesmo verificando os atributos consultados. O histórico de compras realizadas ou informações baixadas são excelentes indícios de avaliação positiva do cliente sobre um produto. Devemos lembrar também que por vezes os clientes deixam sugestões, fornecem avaliações e reclamações que podem ser usadas como insumo para avaliar o produto e também para entender o perfil do cliente.

As informações podem ficar registradas no perfil do cliente caso tenha se identificado no início da interação, ou podem servir como dados anônimos a serem

agregados para representar o perfil médio da comunidade sem as especificidades do indivíduo. As informações da comunidade são úteis. Por exemplo, a lista dos mais vendidos serve para dar uma noção da aceitação de um determinado produto pela comunidade. Essa informação tem impacto na decisão individual.

A saída de um Sistema de Recomendação varia, podendo ser apresentada sob a forma de:

Lista (estatística): lista com os N mais vendidos, mais visitados, mais avaliados

Sugestões: ação pró-ativa de um Sistema de Recomendação de indicar produtos ou informações adequadas a um cliente específico

Avaliações: informação passiva sobre a qualidade de um produto sob o ponto de vista da comunidade como um todo ou sob o ponto de vista de grupos com perfil similar ao do cliente da vez

Resenhas: opiniões individuais que são organizadas para mostrar pontos de vistas específicos de um dado produto ou conteúdo

15.3 Forma de atuação de Sistemas de Recomendação

Os sistemas de recomendação apresentam diferentes comportamentos na interação com o cliente. Pode ter uma **atuação pró-ativa**, envia deliberadamente sugestões aos clientes via correio eletrônico quase como um gerador de ‘spam’. Essa abordagem tem que passar por um consentimento prévio do cliente, em que o cliente se cadastra para receber mensagens sobre as novidades que possam interessar. Funciona como um monitor sobre a disponibilidade de algum produto específico que possa agradar. Vários sites de compras e de leilões oferecem essa funcionalidade. Desta forma, monitores ficam observando as oportunidades de negócio para selecionar as que sejam boas para o cliente.

Sistemas de Recomendação também podem ter uma **atuação sob demanda**. As sugestões são calculadas, mas só são apresentadas se o cliente solicitar. Por exemplo, ao entrar no Amazon.com como cliente identificado (por *login* e senha), é mostrado que existem recomendações para aquele cliente e, para obtê-las, é preciso selecionar o link indicado.

Outro tipo de atuação, que é a mais comum, é a **atuação passiva**. Ao entrarmos num site, encontraremos um conjunto de sugestões disponíveis e contextualizadas pela navegação, por exemplo, haverá sugestões de outros livros quando o cliente estiver na página para comprar um determinado livro, e assim por diante.

Um Sistema de Recomendação também deve se posicionar sobre o grau de personalização da recomendação. Podemos ter **recomendações genéricas**, completamente despersonalizadas que apenas apresentam visões em consequência de comportamento de grupos sobre os produtos, como é o caso da lista de “Mais vendidos” que retrata a preferência no atacado. É uma recomendação mais fácil de ser gerada por consultas em bancos de dados, e requer apenas o registro das operações.

A recomendação pode ser **efêmera**. Por exemplo, ao apresentar correlações de itens comprados em conjunto estamos induzindo a certos gostos comuns. As vendas casadas são inferidas do histórico de compras dos clientes e são fortalecidas pela

frequência com que as vendas casadas são realizadas. A partir da escolha de um produto, outros produtos são sugeridos por serem frequentemente adquiridos em conjunto. É importante observar que a sugestão não é personalizada para o cliente.

Por fim temos a **recomendação persistente**, gerada a partir da análise do comportamento dos usuários e extremamente **personalizada**.

15.4 Métodos de Geração de Recomendações

Recomendações auxiliam pessoas a encontrarem mais rapidamente produtos e conteúdos, o que proporciona maior grau de satisfação. Nas seções anteriores, vimos os diferentes tipos de insumos para gerar uma recomendação, as diferentes formas de apresentar a recomendação, e os graus de atuação e personalização da recomendação. No entanto, como as recomendações são geradas? Como agrupar e interpretar as avaliações dos indivíduos em sugestões consolidadas ou individualizadas?

Existem diversas técnicas de geração de recomendações. As técnicas são classificadas em:

Recomendação baseada em recuperação direta da informação

Recomendação baseada em filtragem colaborativa

Recomendação baseada em filtragem por conteúdo

Recomendação baseada em recuperação direta da informação

O usuário especifica a consulta e o sistema recupera itens que satisfazem à pesquisa realizada. É o método de recomendação mais simples de implementar uma vez que se baseia em consultas diretas nos bancos de dados dos produtos. Para que funcione bem, os produtos ou conteúdos devem estar estruturados e organizados num banco de dados ou no modelo RDFs (*Resource Description Framework*)¹. O método não é muito poderoso, pois a qualidade da recomendação depende da busca no cliente. Também é muito dependente dos critérios de classificação da informação.

A busca direta pode estar associada a uma organização taxonômica do conteúdo, apresentada pelos sites que oferecem produtos e conteúdos. Por exemplo, num site sobre filmes, a organização taxonômica pode ser em função das classes de filmes, como ação e drama. Ao encontrar o nível de detalhamento de categoria desejado, o usuário aciona a solicitação da lista de itens disponíveis.

A implementação deste método também é relativamente simples, e requer a organização hierárquica dos dados, neste caso, apenas sugerindo uma navegação. A qualidade da sugestão depende da qualidade da hierarquia de categorias: se a hierarquia for muito específica, a sugestão provavelmente não atenderá aos objetivos do usuário; se a hierarquia for muito geral, a sugestão requer um esforço cognitivo maior para que o usuário mapeie o que ele quer em uma categoria abstrata disponível.

¹ Modelo padrão para troca de dados na Web preconizado como padrão para a Web semântica para facilitar junção e mapeamento de dados na Web.

As recomendações estatísticas também fazem parte da classe de técnicas de recuperação direta e também representam consultas (*queries*) nas bases de produtos ou conteúdos. “Os mais vendidos”, “Os mais visitados”, “O melhor avaliado” são exemplos de recomendações estatísticas. É preciso associar atributos aos itens para o registro da informação, como por exemplo um contador de visitação. Assim, na apresentação de uma nuvem de itens recomendados, o tamanho da fonte está diretamente relacionado à recomendação: quanto maior, mais recomendado.

Recomendação baseada em Filtragem Colaborativa

Filtragem colaborativa é um método de geração de recomendação que tenta prever o grau de interesse de um cliente em determinados produtos a partir de correlações entre as avaliações feitas por este cliente e as avaliações fornecidas por outros clientes. Avaliações refletem os gostos das pessoas. A hipótese subjacente é que pessoas que avaliaram um grande conjunto de produtos de maneira semelhante, pelo menos num futuro próximo devem continuar avaliando de maneira semelhante novos produtos. Desta forma, se quisermos sugerir um produto a um cliente ou prever como esse cliente avaliará um produto, devemos utilizar os dados de uma pessoa que venha avaliando ao longo do tempo de maneira semelhante a este cliente. O novo produto avaliado por um, deve receber avaliação semelhante pelo outro.

O método consiste no seguinte conjunto de passos:

1. Calcular a similaridade entre os usuários. Em Filtragem Colaborativa, usa-se a avaliação dada pelos usuários aos produtos como o parâmetro para calcular a similaridade entre os usuários. Distância entre dois usuários = 1-correlação das avaliações desses usuários. Podemos usar qualquer técnica de correlação, como por exemplo, a de Pearson (Coeficiente de Pearson).

A correlação mede a distância entre avaliações calculadas. Cada coluna representa um usuário distinto e cada linha representa um produto distinto. A célula da tabela contém a avaliação que um usuário u deu a um produto p . A distancia mede a média das avaliações entre dois usuários i e j a partir do conjunto de avaliações que deram sobre os mesmos produtos.

2. Selecionar os vizinhos mais próximos. Fica a critério do desenvolvedor que escolher qual a distância entre os vizinhos.

3. Fazer previsão sobre avaliação do usuário alvo para um produto não consumido. Através da previsão (ou predição) calcularemos o quão válido ou útil será o produto para o usuário alvo. Algo como “tentaremos adivinhar a avaliação que o usuário daria ao produto”. Tal previsão, P , se dá pela seguinte formula:

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

Em que:

$P_{u,i}$ representa a previsão de avaliação do item i pelo usuário u

\bar{r}_u representa a avaliação (*rating*) média do usuário u considerando todos os produtos que ele avaliou

\bar{r}_j avaliação média que o produto j considerando todas as avaliações que foram feitas para tal produto

$w_{u,j}$ representa a similaridade entre o usuário u e o usuário j

$r_{j,i}$ representa a avaliação para o produto alvo i feita pelo usuário j

$r_{j,i} - \bar{r}_j$ representa a diferença entre a avaliação do produto alvo pelo j e a avaliação média para o filme alvo

A fórmula acima retrata a atualização da expectativa sobre a avaliação do cliente a um novo produto tendo em vista a avaliação dos clientes considerados pares a tal produto. Sem considerar a filtragem colaborativa, tal expectativa é a avaliação média que o cliente deu aos produtos já avaliados por ele, aqui retratada como \bar{r}_u . A atualização desta expectativa está na parcela acrescida ao \bar{r}_u . Essa parcela representa a média, ponderada pela distância entre o cliente em foco e cada um de seus “ n ” similares, das diferenças entre as avaliações dadas ao produto em questão e avaliações médias de cada avaliador considerado similar. O número de clientes considerados similares “ n ” é geralmente definido pelo dono do negócio.

Filtragem colaborativa é um método eficiente por seu poder discriminatório. Pesquisas revelam que as recomendações são relevantes e a qualidade da recomendação melhora com o aumento do número de avaliações e de usuários no processo. Embora seja muito utilizado, vale ressaltar que é um método computacionalmente caro, pois a matriz de similaridade deve ser calculada (leva muito tempo para gerá-la) e armazenada para possibilitar respostas rápidas ao uso. Além da alta complexidade computacional, o método não trabalha bem com matrizes esparsas: para matrizes com poucos usuários e avaliações, o método tende a dar resultados ruins. O método foi pensado para ser usado por grandes grupos ativos que voluntariamente avaliam produtos para alimentar o processo. Outra desvantagem é que o método requer que os produtos sejam avaliados para entrar no processo e desta forma os lançamentos não aparecem nas recomendações.

Recomendação baseada em Filtragem por Conteúdo

Método de recomendação baseado nas informações sobre o conteúdo dos itens. A Filtragem por Conteúdo usa algoritmos de aprendizagem de máquina para induzir um perfil das preferências de um usuário a partir de exemplos, tendo em vista uma descrição das características dos conteúdos. Por exemplo, filmes podem ser categorizados como ação, terror, romance ou comédia. O filme é classificado a partir da maioria das características que se enquadram numa determinada categoria.

O método é composto de 4 passos:

Classifique os itens avaliados segundo as categorias pré-estabelecidas

Para cada categoria, calcule a avaliação média de cada avaliador

Ordene os itens avaliados para obter listas de preferências ordenadas por categorias

Calcule a previsão de avaliação segundo a fórmula apresentada a seguir. A previsão da utilidade do produto “i” para o usuário “u” é calculada como a avaliação média do produto “i” ajustada pela avaliação média (normalizada) demonstrada pelo usuário “u” para os produtos classificados na classe “categoria”

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

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

$\bar{r}_i \rightarrow$ “Avaliação_Média_do_Item i”

$\bar{r}_{categoria,u} \rightarrow$ “Avaliação_Média_do_usuario_U da_Categoria_da_qual_o_item_pertence”

O método de filtragem por conteúdo apresenta as seguintes vantagens:

Não necessita de dados de outros usuários

É capaz de recomendar itens novos ou não populares

Sem problemas de Partida a Frio ou de Esparsidade, que são problemas que ocorrem quando um novo item é adicionado ao conjunto e não contém avaliações prévias

Permite elaboração de recomendações a usuários com gostos exclusivos

Pode fornecer explicações sobre os itens recomendados listando as características do conteúdo que conduziram à recomendação

É uma tecnologia madura

O método apresenta as seguintes desvantagens:

As recomendações são estáticas, desprezam a opinião dos usuários

O conteúdo precisa ser codificado por características inteligíveis porque muitas vezes o conteúdo é analisado automaticamente para se identificar as categorias

Possui baixa eficiência se o conteúdo não for muito informativo

Para livros técnicos o título é relevante pois representa a categoria, mas para novelas ou filmes, o título pode trazer pouca informação pois não possui uma relação estreita para indicar uma categoria. Com isso fica mais difícil gerar as categorias.

Recomendação baseada em Descoberta de Conhecimento

Muitas vezes a recomendação é baseada em relações sobre a interação “Produto e Clientes”. A recomendação não é baseada diretamente em avaliações, mas sim na frequência com que as ações acontecem conjuntamente. Por exemplo, se um conjunto de clientes compra ao mesmo tempo 2 produtos X e Y, quando alguém comprar X pode-se esperar que também compre Y. Essa recomendação é gerada com base na técnica vinda da área de mineração de dados, mais especificamente a técnica de regras de associação (Agrawal, 1993).

Veamos como o método funciona. Num site de vendas, considere que cada interação do cliente forme uma transação que pode ser visto como um conjunto de itens. Uma regra de associação $X \rightarrow Y$, significa que quando se encontra o comportamento X é

esperado que o comportamento Y aconteça. X e Y podem ser ações únicas ou conjunto de ações. Uma ação pode ser uma avaliação, uma compra ou simplesmente uma navegação.

A regra $X \rightarrow Y$ possui dois critérios de geração: suporte e confiança. Suporte representa a frequência com que X e Y aparecem juntos na base de dados, enquanto a confiança representa a probabilidade de Y aparecer uma vez que X aconteceu. Para exemplificar os conceitos, analisaremos a pequena base apresentada na Tabela 15.1.

Tabela 15.1: Avaliações de Filmes

ID	Hannibal	Bad Boys	Comer, Rezar, Amar	A pequenina	Batman	A Era do Gelo III	Up
1	✓	✓		✓			
2	✓	✗	✓	✗	✓	✗	
3	✗	✓		✓			
4	✓		✓	✓	✗	✓	✓
5	✗			✗			
6		✓	✗	✓	✓		
7	✓	✗		✗	✗	✓	✓
8			✗	✗	✓		
9	✓	✗		✓	✗	✓	✓
10	✗	✓		✓	✗		

(Legenda: ✓ = Gostei ✗ = Não-gostei)

É interessante notar que nossa intuição nos levaria a uma expectativa errada de que os amantes do filme "Bad Boys" também gostem do filme "A pequenina" e vice-versa, porém a técnica nos ajuda a entender que tais expectativas deveriam ser diferentes. Para esse cálculo, é desejado um suporte superior a 30% e uma confiança superior a 70%. Isso significa que há interesse em regras que reflitam padrões que apareçam pelo menos em 30% das transações e que dada a premissa possa-se criar uma expectativa de pelo menos 70% que a conclusão se seguirá. Para o exemplo ilustrado, calculando suporte e confiança, chegamos a:

$$F_{sup}^{BadBoys, Pequenina} = \frac{4}{10} = 40\%$$

$$F_{confiança}^{BadBoys \rightarrow Pequenina} = \frac{40\%}{40\%} = 100\%$$

$$F_{confiança}^{Pequenina \rightarrow BadBoys} = \frac{40\%}{60\%} = 66\%$$

É interessante notar que se espera que os amantes do filme “Bad Boys” também gostem do filme “A pequenina”, mas não vice-versa. O método de geração de recomendação a partir de regras de associação é bastante utilizado. Por exemplo, após a seleção de um livro na Amazon é apresentada a famosa recomendação: “Pessoas que compraram este livro, também compraram esses outros itens aqui: <lista de produtos>”.

Dentre as vantagens do método destacamos que a recomendação é gerada pela correlação entre itens, o que pode ser calculado a priori. No entanto o processamento é computacionalmente caro e cresce exponencialmente com o aumento do número de produtos.

15.5. Desenvolvimento de um projeto de Sistema de Recomendação

Para o desenvolvimento de um sistema de recomendação que trabalhará embutido em uma aplicação, devem ser considerados os objetivos específicos que se quer alcançar. São apresentados, a seguir, cenários e o tipo de Sistema de Recomendação de menor complexidade que se ajusta ao objetivo relatado.

Cenário 1: Listas de Recomendação Baseadas em Buscas

Neste cenário, o objetivo é atender aos usuários cujas necessidades são definidas por eles próprios. Espera-se que o usuário forneça para o sistema as palavras chave ou atributos referentes à busca desejada para ter acesso às recomendações. A apresentação da recomendação é feita sob demanda e existe pouca personalização dos produtos e serviços oferecidos.

Cenário 2: Avaliações e Comentários de Consumidores

O objetivo é obter a credibilidade do serviço de recomendação por meio da opinião da comunidade. O principal alvo são as vendas a varejo no comércio eletrônico, onde há necessidade de exorcizar o fantasma da baixa credibilidade. Sistemas de recomendação podem utilizar a navegação do usuário para sugerir a leitura de revisões, comentários, avaliações e previsões – as opiniões da comunidade são bem aceitas pelos usuários. A entrega da recomendação é passiva e praticamente não há personalização.

Cenário 3: Recomendações Associadas a Produtos

O objetivo do negócio é fazer as “vendas casadas”, propor ao usuário a compra de itens que de alguma forma estão relacionados a outro que ele esteja adquirindo. Utilizando a navegação do cliente como indicação de interesse, sistemas empregam correlação item a item a partir do histórico de compras da comunidade. A escolha do que recomendar é um caso clássico de Mineração de Dados. A entrega da recomendação é feita de forma passiva. A personalização de produtos ou serviços é efêmera.

Cenário 4: Personalização Profunda (Influência do Domínio de Produtos)

O objetivo é recomendar, dentro de um domínio específico, com mais precisão e foco no usuário. O domínio influencia a escolha e, para verificar a necessidade de sistemas de recomendações, é necessário considerar: (i) número de itens em um site: quando o número é pequeno, um sistema de recomendação é pouco útil; (ii) itens de elevado preço: quando a aquisição de um item for algo que exija um estudo financeiro, é interessante dispor de sistemas de recomendação; (iii) intervalo de tempo entre atualizações de catálogo: catálogos estáveis dão tranquilidade aos usuários, enquanto catálogos voláteis requerem o uso de Sistema de Recomendação; (iv) homogeneidade de produtos no domínio: quando a variedade de itens semelhantes é muito grande, os Sistema de Recomendação são uma ajuda valiosa; (v) proporção de produtos relacionados por atributos específicos: a seleção por atributos específicos exige pouco dos Sistema de Recomendação, que são tão mais úteis quanto menos o consumidor tiver certeza do que quer adquirir. O tipo de entrega da recomendação é ativo e a existe personalização de produtos ou serviços. A Tabela 15.2 resume os cenários descritos.

Tabela 15.2 – Resumo dos cenários de desenvolvimento de Sistema de Recomendação

	Listas de Recomendação Baseadas em Buscas	Avaliações e Comentários de Consumidores	Recomendações Associadas a Produtos	Personalização Profunda
Objetivo:	Atender usuários com necessidades definidas	Obter credibilidade através da opinião da comunidade	Fazer vendas “casadas”	Recomendar dentro de um domínio específico
Descrição:	O usuário fornece uma palavra chave ou atributo para ter acesso a recomendações	Vendas a varejo	Empregam correlação item a item e o histórico de compras comunidade	O domínio influencia a escolha do sistemas de recomendações e é necessário considerar uma série de questões
Tipo de entrega:	Sob demanda ou “pull”	Passiva	Passiva	Ativa
Grau de personalização	Personalização efêmera de produtos ou serviços	Não há personalização	Personalização efêmera de produtos ou serviços	Há personalização de produtos ou serviços

Experimente diferentes sistemas de recomendação

Para você sentir o poder de um sistema de recomendação de produtos e informações, inscreva-se em sites que utilizem esta técnica. Por exemplo, inscreva-se no site MovieLens que tem por objetivo sugerir filmes. Este site é de um dos pioneiros na área de Sistema de Recomendação e implementa a técnica de filtragem colaborativa. Acesse o endereço <<http://movielens.umn.edu>> e registre-se criando um login e senha. Ao entrar na página de criação de uma conta, além das informações tradicionais sobre dados pessoais, você deverá avaliar um conjunto de 15 filmes para que o sistema comece a entender seus gostos. Faça isso e depois passeie pelas páginas e analise as recomendações exibidas para vocês. Outra sugestão é que você entre em outros sites conhecidos que possuem sistemas de recomendação acoplados como:

www.Amazon.com

www.Submarino.com.br

www.LivrariaSaraiva.com.br

www.americanas.com.br

www.ebay.com

Leitura recomendada

Reategui, E.B. e Cazella, S.C.- Sistemas de Recomendação, V ENIA – XXV Congresso da Sociedade Brasileira de Computação, 2005

Referências

Resnick, P.; Iacovou, N.; Sushak, M.; Bergstrom, P.; Riedl, J. GroupLens: An open architecture for collaborative filtering of netnews. ACM Conference on Computer Supported Collaborative Work Conference, 10/1994, Chapel Hill, NC, p.175-186, 1994.

Riedl, J.; Konstan, J.; Vrooman, E. Word of Mouse: The Marketing Power of Collaborative Filtering, Business Plus, 2002.

Segaran, T. Programming Collective Intelligence: Building Smart Web 2.0 Applications, O'Reilly Media, 2007.

Marmanis, H.; Babenko, D. Algorithms of the Intelligent Web, Manning Publications, 2009.

Exercícios

15.1 Considere que a locadora “Alvo” possui um conjunto de filmes para alugar e que tais filmes recebem avaliações de seus clientes. Atualmente consta com o seguinte conjunto de filmes e avaliações apresentados na Tabela 15.A.1 (disponível no site do livro). José é um excelente cliente da locadora Alvo. Ele vai ficar em casa sábado e deseja uma sugestão de filme para alugar. Ele vai à locadora e observa que acaba de

chegar o filme “Comer, Rezar e Amar”. Utilizando a técnica de filtragem colaborativa, usando a correlação de Pearson, verifique se ele vai gostar do filme. Considere verdadeira a afirmativa de que “ele vai gostar” se a expectativa de avaliação dele for superior a 3. Justifique sua resposta mostrando o passo a passo do método. No seu cálculo considere apenas os 3 clientes mais similares a José.

15.2 Usamos o mesmo exemplo dos filmes da locadora Alvo para explicar o método de filtragem por conteúdo. Para tanto, o conjunto de filmes teve que ser organizado segundo a seguinte categoria: Ação, Drama, Comédia, Romance e Terror, apresentado na Tabela 15.A.4 (disponível no site do livro). Qual a expectativa de avaliação de José para o filme “Comer, Rezar, Amar”?

15.3-Quais são os principais tipos de sistemas de recomendação?

Quais as diferenças entre eles?

Quando você considera cada um pode /deve ser usado

Respostas comentadas

15.1 A Tabela a seguir apresenta alguns cálculos intermediários para que entendamos o cálculo da previsão da avaliação do José tendo em vista a avaliação dos seus pares em avaliação de filmes.

Inicialmente devemos calcular a matriz de distância entre os clientes. Usando a correlação de Pearson, chegamos a tabela 15.3:

Tabela 15.3. Distância entre os usuários da locadora Alvo.

	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

Com isso, chegamos a conclusão que Jaime, Manoel e Maria são os 3 clientes com avaliações mais similares a Jose.

Para calcularmos a previsão da avaliação de José para o filme “Comer, Rezar, Amar”, devemos levar em conta as avaliações dadas ao filme por esses 3 clientes considerados mais similares a Jose.

Logo, podemos esperar que a avaliação do filme “Comer, Rezar, Amar” pelo José seja calculada como: $3,1 + (-0,497/1,99)$, que é igual a 2,85. Logo, podemos considerar que ele NAO vai gostar do filme. Interessante notar que se considerássemos apenas a media das avaliações de José, criaríamos uma expectativa diferente: a de que ele iria gostar do filme.

15.2 Calculando a avaliação média por categoria, por usuário, chegamos aos resultados apresentados na Tabela 15.4.

Tabela 15.4. Media de avaliações por categoria, por usuário, da locadora Alvo.

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

O próximo passo será calcularmos a média das avaliações por categoria, como apresentada na Tabela 15.5.

Tabela 15.5. Media de avaliações por categoria.

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

Com isso, podemos calcular a expectativa para a avaliação de Jose aos novos filmes que entraram na locadora, inclusive o “Comer, Rezar e Amar”, como apresentado na Tabela 15.6. Como mostrado na tabela, também chegaríamos à conclusão que Jose NÃO vai gostar do filme.

Tabela 15.6. Expectativa de avaliação de filmes novos que entraram na locadora Alvo, pelo José.

Novos Filmes	Classe	Avaliação Média da Categoria	Avaliação média do Filme	Aval média do José por Categoria	Avaliação José para Filme
Comer, Rezar, Amar	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

Ana Cristina Bicharra Garcia



<http://lattes.cnpq.br/4879977915136752>

Graduada em Engenharia Civil pela Universidade Federal do Rio de Janeiro (1983), Mestrado em Computer Aided Civil Engineering - Stanford University (1989) e Doutorado em Computer Aided Civil Engineering - Stanford University (1992). Atualmente, é Professora Titular do departamento de Ciência da Computação da Universidade Federal Fluminense (UFF) e pesquisadora 2 do CNPq. Também é Coordenadora do Laboratório de Documentação Ativa e Design Inteligente (ADDLabs), laboratório da UFF em parceria com a Petrobras. Atua como colaboradora em projeto europeu, MADRINET, coordenado pela Universidad Carlos III de Madrid e consultora ad-hoc de órgãos de fomento à pesquisa, como CNPq, CAPES e FAPERJ. As atuais áreas de interesse incluem Inteligência Artificial, mais especificamente documentação ativa, aquisição de conhecimento, mineração de texto e sistemas multiagentes, e interação homem-máquina, mais especificamente e-gov e trabalho em grupo.

Claudia Lage Rebello da Motta



<http://lattes.cnpq.br/0774464575739440>

Claudia Lage Rebello da Motta possui graduação em Informática pela Universidade Federal do Rio de Janeiro (1984), mestrado (1989) e Doutorado (1999) em Engenharia de Sistemas e Computação pela COPPE - Universidade Federal do Rio de Janeiro. É analista de TI da

Universidade Federal do Rio de Janeiro e faz parte do corpo docente do Programa de Pós-Graduação em Informática (PPGI/UFRJ). Diretora do Instituto Tércio Pacitti de Aplicações e Pesquisas Computacionais da Universidade Federal do Rio de Janeiro (iNCE/UFRJ). Tem experiência na área de Ciência da Computação, com ênfase em Informática Aplicada à Educação, atuando principalmente nos seguintes temas: Ambientes Educacionais Cooperativos, Construção da Inteligência Coletiva e Neuropedagogia e Informática.

Flávia Maria Santoro



<http://lattes.cnpq.br/5377746284077362>

Flávia Maria Santoro possui graduação em Engenharia Eletrônica pela Universidade Federal do Rio de Janeiro (1989), mestrado (1996) e Doutorado (2001) em Engenharia de Sistemas e Computação pela COPPE - Universidade Federal do Rio de Janeiro. É professora adjunta e membro do Programa de Pós-graduação em Informática da Universidade Federal do Estado do Rio de Janeiro - UNIRIO. Coordena o Núcleo de Pesquisa e Prática em Tecnologia (NP2Tec). Tem experiência na área de Ciência da Computação, com ênfase em Sistemas de Informação, atuando principalmente nos seguintes temas: Gestão de Processos de Negócios Gestão de Conhecimento e Aprendizagem Organizacional, Trabalho Colaborativo Apoiado por Computador.

Adriana S. Vivacqua



<http://lattes.cnpq.br/6494676052801758>

Adriana Santarosa Vivacqua possui graduação em Tecnologia em Processamento de Dados pela Pontifícia Universidade Católica do Rio de Janeiro (1993), mestrado em Computação pela Universidade Federal Fluminense (1997), mestrado em Media Arts and Sciences - Massachusetts Institute of Technology (1999) e doutorado em Engenharia de Sistemas e Computação pela Universidade Federal do Rio de Janeiro (2007) e Université de Technologie de Compiègne. Atualmente é professora do Departamento de Ciência da Computação da Universidade Federal do Rio de Janeiro e pesquisadora colaboradora no ADDLabs/UFF. Atua na área de Interação Homem Computador e Trabalho Cooperativo Apoiado por Computador, atuando principalmente nos seguintes temas: trabalho cooperativo apoiado por computador, agentes inteligentes, interfaces com usuário e personalização e contexto em recuperação de informação. Atualmente coordena o projeto Acompanhamento de Planos de Emergências (APE), financiado pela FAPERJ.



Jonice Oliveira

<http://lattes.cnpq.br/0990344839864230>

Jonice Oliveira obteve o título de Bacharel em Informática (2001) pela Universidade Federal do Rio de Janeiro e o de Mestre em Engenharia de Sistemas e Computação (2003) pela

COPPE/UFRJ. Concluiu o Doutorado em Engenharia de Sistemas e Computação na COPPE/UFRJ (2007). Na mesma instituição realizou o seu Pós-Doutorado, concluindo-o em 2008. Atualmente é professora adjunta da Universidade Federal do Rio de Janeiro (UFRJ) e suas principais áreas de pesquisa são Gestão do Conhecimento, Análise de Redes Sociais, Computação Autônoma e Web Semântica. Durante o seu doutorado recebeu o prêmio IBM Ph.D. Fellowship Award.

Anexo Digital

Exercício 15.1

Considere que a locadora “Alvo” possui um conjunto de filmes para alugar e que tais filmes recebem avaliações de seus clientes. Atualmente conta com o seguinte conjunto de filmes e avaliações apresentados na Tabela 15.A.1.

Tabela 15.A.1. Banco de dados de avaliações de filmes da Locadora Alvo.

	Mar ia	Jos é	Joã o	Ped ro	Lau ra	Man oel	Jai me	Rafa el	Luc as	Reinal do	Feli pe	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	2	2	4	3	4	3	2	4	5	3	4	4

(2003)												
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
Read Review	5	3	2	4	1	4	4	3	3	3	2	1
Surviving Picasso	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	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	3	3	1	4	2	4	3	1	1	5	2	1

(2003)												
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

Usando a correlação de Pearson, chegamos ao conjunto de correlações entre usuários apresentadas na Tabela 15.A.2.

Tabela 15.A.2. Correlações entre as avaliações dos clientes da locadora Alvo.

	Maria	José	João	Pedro	Laura	Manoel	Jaime	Rafael	Lucas	Reinaldo	Felipe	Eva
Maria	1,00	0,59	-0,40	0,68	-0,70	0,55	0,67	-0,52	-0,41	0,72	-0,75	-0,73
José		1,00	-0,46	0,47	-0,35	0,64	0,76	-0,51	-0,44	0,55	-0,45	-0,34
João			1,00	-0,51	0,60	-0,45	-0,56	0,77	0,77	-0,58	0,58	0,59
Pedro				1,00	-0,63	0,62	0,66	-0,70	-0,49	0,78	-0,66	-0,67

Laura					1,00	-0,34	-0,41	0,56	0,57	-0,66	0,88	0,87
Manoel						1,00	0,66	-0,54	-0,40	0,59	-0,29	-0,36
Jaime							1,00	-0,64	-0,45	0,56	-0,45	-0,47
Rafael								1,00	0,72	-0,63	0,56	0,63
Lucas									1,00	-0,46	0,58	0,56
Reinaldo										1,00	-0,73	-0,68
Felipe											1,00	0,85
Eva												1,00

Tabela 15.A.3. Distancia entre os usuários da locadora Alvo.

	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

Considerando que a distância é calculada como “1- correlação”, chegamos às distancias entre os usuários que são apresentados na Tabela 15.A.3.

Tabela 15.A.3. Distancia entre os usuários da locadora Alvo.

	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

Se desejássemos saber os três usuários mais similares a José, deveríamos escolher Jaime, Manoel e Maria, nesta ordem. Assuma que entrou um novo filme na Praça: “Comer, Rezar, Amar” e que Jaime avaliou com 2, Manoel com 3 e Maria com 4 estrelas. Qual deverá ser a avaliação de José? Será que seria recomendável ele ir assistir?

Exercício 15.2.

Usamos o mesmo exemplo dos filmes da locadora Alvo para explicar o método de filtragem por conteúdo. Para tanto, o conjunto de filmes teve que ser organizado segundo a seguinte categoria: Ação, Drama, Comédia, Romance e Terror, apresentado na Tabela 15.A.4.

Tabela 15.A.4. Filmes da locadora Alvo, avaliados e organizados por categorias.

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
Media	3,70	4,30	2,60	3,40	3,40	4,40	4,40	2,40	2,90	3,70	3,40	3,40

Comédia	Maria	José	João	Pedro	Laura	Manoel	Jaimé	Rafael	Luca	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
Media	2,44	2,56	4,33	2,44	3,67	2,67	2,33	4,44	4,56	2,56	3,78	3,78

Drama	Maria	José	João	Pedro	Laura	Manoel	Jaimé	Rafael	Luca	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

Arte	Maria	José	João	Pedro	Laura	Manoel	Jaimé	Rafael	Luca	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

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

Calculando a avaliação media por categoria, por usuário, chegamos aos resultados apresentados na Tabela 15.A.5.

Vejam os que acontecem quando um conjunto de novos filmes entra na locadora, incluindo o “Comer, Rezar, Amar”, recebendo as avaliações apresentadas na Tabela 15.A.7.

Tabela 15.A.7. Avaliação de novos filmes da locadora Alvo.

Novos Filmes	Classe	Pedro	Laura	Manoel	Jaime	Rafael	Lucas	Reinaldo	Felipe	Eva	Avaliação Média do Filme
Comer, Rezar, Amar	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

Qual a expectativa de avaliação de José para o filme “Comer, Rezar, Amar”?

